

## A PRINCIPAL COMPONENT ANALYSIS OF THE DIFFUSE INTERSTELLAR BANDS

T. ENSOR,<sup>1</sup> J. CAMI,<sup>1,2</sup> N.H. BHATT,<sup>1</sup> AND A. SODDU<sup>1</sup>

<sup>1</sup>*Department of Physics and Astronomy and Centre for Planetary Science and Exploration (CPSX), The University of Western Ontario, London, ON N6A 3K7, Canada*

<sup>2</sup>*SETI Institute, 189 Bernardo Ave, Suite 100, Mountain View, CA 94043, USA*

### ABSTRACT

We present a principal component analysis of 23 line of sight parameters (including the strengths of 16 diffuse interstellar bands, DIBs) for a well-chosen sample of single-cloud sightlines representing a broad range of environmental conditions. Our analysis indicates that the majority ( $\sim 93\%$ ) of the variations in the measurements can be captured by only four parameters. The main driver (i.e., the first principal component) is the amount of DIB-producing material in the line of sight, a quantity that is extremely well traced by the equivalent width of the  $\lambda 5797$  DIB. The second principal component is the amount of UV radiation, which correlates well with the  $\lambda 5797/\lambda 5780$  DIB strength ratio. The remaining two principal components are more difficult to interpret, but are likely related to the properties of dust in the line of sight (e.g., the gas-to-dust ratio). With our PCA results, the DIBs can then be used to estimate these line of sight parameters.

*Keywords:* ISM:lines and bands — ISM:molecules — methods:data analysis — methods:statistical

## 1. INTRODUCTION

One of the greatest outstanding astronomical challenges is the identification of the diffuse interstellar bands (DIBs): a series of  $\sim 500$  absorption features detected in optical and infrared spectra toward reddened stars (see [Herbig 1995](#); [Sarre 2006](#); [Snow 2014](#) for reviews and [Hobbs et al. 2008, 2009](#) for recent surveys). It has been clear that the DIBs arise from material in interstellar clouds since they were first detected ([Heger 1922](#)); but despite nearly 100 years of research, most of the DIB carriers remain unidentified. The only notable exception is the recent identification of four DIBs as due to  $C_{60}^+$  ([Campbell et al. 2015](#); [Walker et al. 2015](#)). This identification is in line with the general consensus that the DIB carriers are highly stable, carbonaceous, gas-phase molecules.

An identification of DIBs with specific carriers requires a perfect match between laboratory spectra and astronomical observations; however, given the countless numbers of possible carrier candidates, this is not an easy task. To guide these laboratory efforts, observational studies aim to learn about the nature of the carriers and constrain the set of possible species. Two types of such studies that are particularly relevant for this paper are correlation studies – either mutual DIB correlations, or correlations between the DIBs and line of sight properties (see e.g. [Seab & Snow 1984](#); [Herbig 1993](#); [Cami et al. 1997](#); [McCall et al. 2010](#); [Friedman et al. 2011](#)) – and research into the environmental behavior of the DIBs (e.g. [Jenniskens et al. 1994](#); [Cami et al. 1997](#); [Sonnentrucker et al. 1997](#); [Cox et al. 2006](#)).

The basic idea behind pairwise correlations is simple: If two DIBs arise from the same state in the same carrier, they should have the same strength ratio in all lines of sight and thus, their equivalent widths (EWs) should exhibit a perfect correlation. Observational studies have not found two DIBs that show such a perfect correlation. The best case is the  $\lambda 6196$  and  $\lambda 6614$  DIBs which correlate well in a large sample of sightlines (correlation coefficient  $r$  of 0.986; see [McCall et al. 2010](#)). However, even these two DIBs show quite different behavior in the remarkable sightline towards Herschel 36 implying that they are most likely originating from different carriers ([Dahlstrom et al. 2013](#); [Oka et al. 2013](#)). This has led to the “one DIB, one carrier” paradigm ([Herbig 1995](#); [Cami et al. 1997](#); [Snow 2014](#)). At the same time, the notion of DIB “families” can be established: sets of DIBs that correlate fairly well with one another and that might have similar or chemically related carriers ([Krelowski & Walker 1987](#); [Cami et al. 1997](#)). There are two important caveats though in correlation studies. First, correlation studies generally include only a

small number of DIBs, typically fairly strong and narrow DIBs. Second, while the role of measurement uncertainties on the correlation coefficient is well established (see e.g. discussions in [Herbig 1975](#); [Cami et al. 1997](#)), they are often not taken into account in correlation studies.

Correlations between the DIB strengths and line of sight parameters can reveal additional properties about the DIB carriers. DIB strengths often show some correlation ( $r$  typically 0.7) with various other line of sight parameters, albeit typically with a large scatter around the mean relation; examples are the correlation with  $E(B - V)$ , or the  $\lambda 5780$  DIB strength with  $N(\text{H I})$  or the column densities of other interstellar species (see e.g. [Herbig 1975, 1993, 1995](#); [Krelowski et al. 1999](#); [Welty et al. 2006](#); [Friedman et al. 2011](#); [Lan et al. 2015](#); [Baron et al. 2015](#)). A particularly intriguing finding is a subset of DIBs (the so-called “ $C_2$ -DIBs”) that roughly correlate with  $N(C_2)$  and are thought to be chemically related to  $C_2$  or else form under similar conditions ([Thorburn et al. 2003](#)). Modern surveys have confirmed such relations for averaged DIB strengths on large scales, and have furthermore also shown that much of the scatter can be traced back to differences in the amounts of  $H_2$  relative to  $H I$  in the line of sight ([Herbig 1993](#); [Lan et al. 2015](#)).

Part of the scatter in these correlations must thus be due to changes in the physical environment that drive the carrier abundances. Indeed, the DIBs exhibit clear environmental behavior, and show intensity variations that could be explained for instance by ionization or (de-)hydrogenation ([Jenniskens et al. 1994](#); [Cami et al. 1997](#); [Sonnentrucker et al. 1997](#)). Interestingly, various parameters have been shown to be indicative of these environmental conditions. The strength ratio between two strong DIBs,  $\lambda 5797$  and  $\lambda 5780$ , is highly variable and a good indicator of local conditions ([Krelowski et al. 1997](#)). Using this ratio, diffuse clouds can typically be subcategorized into two groups –  $\sigma$  and  $\zeta$  type clouds, named after their prototypes  $\sigma$  Sco (HD 147165) and  $\zeta$  Oph (HD 149757).  $\sigma$  clouds have lower  $W(\lambda 5797)/W(\lambda 5780)$  ratios and are characterized by stronger UV exposure;  $\zeta$  clouds, on the other hand, probe deeper layers of diffuse clouds where material is sheltered from UV radiation, and this causes a much larger  $W(\lambda 5797)/W(\lambda 5780)$  ratio while simultaneously affecting the dust properties ([Cami et al. 1997](#)).

The large number of DIBs coupled with the lack of strong correlations suggest that the DIBs carry an enormous diagnostic potential to study the environments in which they reside. At the same time, it raises the question of what factors drive these variations in the DIB strengths, and how it is possible that there is such a

lack of correlations in such a large collection of spectral lines. Here, we address these questions, and in particular the key question: *how many parameters do we need to explain the variations in the DIB spectrum and what are those parameters?*

To this end, we present a multivariate analysis of a set of strong and clean DIBs with several line of sight parameters. In a proof-of-concept study, we first perform a principal component analysis (PCA) on the data to find out how many parameters are required to describe the observed variations among the DIBs. We physically interpret these new parameters and find convenient quantitative alternatives to represent these parameters. From this work, the huge diagnostic potential of the DIBs becomes clear: since DIBs are products of their environments, we can use DIBs to determine physical parameters of their environment – even without identifying the carriers.

This paper is organized as follows. In Section 2, we describe our sample selection and how we acquired our data. We then describe PCA in Section 3, followed by our results in Section 4. We interpret the results in Section 5 and present our conclusions in Section 6.

## 2. DATA, OBSERVATIONS & METHODS

### 2.1. Target Selection

Our goal in this paper is to determine the parameters that drive the variations in the DIB spectrum. Thus, we need to select a sample of sightlines where physical conditions are reasonably well determined, and that represent the overall observed variations of the DIB spectrum. The first requirement implies that we should restrict ourselves as much as possible to single-cloud lines of sight to avoid having to deal with ill-defined averages throughout multiple intervening clouds. The second requirement stresses the importance of including lines of sight that are as observationally different as possible. Finally, in order to be able to relate any changes to known observables, we need to pick lines of sight for which auxiliary data (e.g. hydrogen column densities,  $E(B - V)$ , extinction properties, ...) are known and available. In this pilot study, we chose to restrict ourselves to include only a limited number of DIBs, and to lines of sight for which we can find high-resolution spectra that allow us to exclude possible blends with stellar lines.

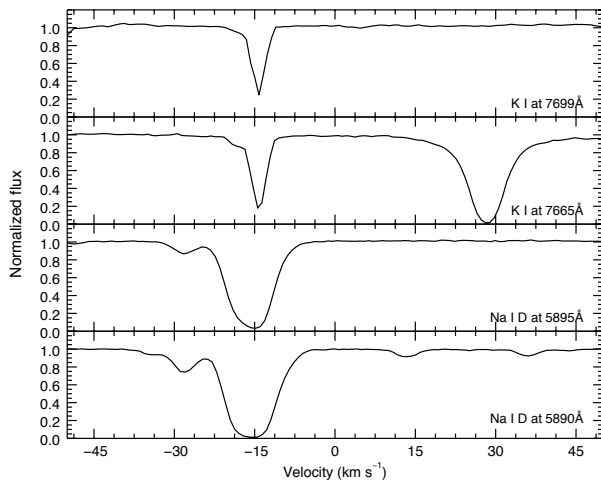
We started our selection of targets from the thorough and detailed study of elemental depletion in the lines of sight towards 243 stars published by Jenkins (2009). This study critically reviews available literature data for  $E(B - V)$ ,  $N(\text{H I})$ ,  $N(\text{H}_2)$ ,  $N(\text{H})$ , and introduces a depletion strength factor,  $F_*$ , describing the collective level of elemental depletion in a line of sight. Starting from

this sample thus ensures a consistent treatment of the required auxiliary line-of-sight data and allows us to ensure a wide coverage of environmental conditions (to the extent that they can be traced by any of these parameters).

We searched the VLT/UVES and ELODIE archives to find good-quality, high-resolution spectra of these targets. UVES (the Ultraviolet and Visual Echelle Spectrograph) is a high-resolution instrument on the VLT covering wavelength ranges from 3000 - 4000Å and 4200 - 11,000Å, with maximum resolutions of 80,000 and 110,000, respectively (Dekker et al. 2000). ELODIE is an echelle spectrograph on the 1.93m telescope at the Observatoire de Haute-Provence in France. ELODIE has a resolution of 42,000 and covers the wavelength range from 3906 - 6801Å (Moultaka et al. 2004). We found appropriate data for 91 of the Jenkins targets: 43 targets in the UVES database; the remaining 48 from the ELODIE database. In a few rare cases, parts of the spectrum would be of too low quality, or simply missing from the data, and in those cases we supplemented our data with archival spectra from the ESPaDOns (Echelle spectropolarimetric device for the observation of stars) instrument on the Canada-France-Hawaii Telescope (CFHT) – a bimodal instrument with maximum resolutions of 68,000 and 81,000 for its spectropolarimetric and non-polarimetric modes, respectively, covering a wavelength range of 3700 - 10,000Å (Donati 2003).

To further select only single-cloud lines of sight, we examined the interstellar Na I D lines at 5890 and 5895Å and the K I lines at 7665 and 7699Å (see e.g., Bhatt & Cami (2015), illustrated in Figure 1). For our current study, we consider a sightline to be a single cloud if these interstellar lines show only one dominant component at the spectral resolution of UVES or ELODIE. Thus, a target was still considered to be a single cloud if there were multiple radial velocity components, but one component had significantly stronger features than the others. For example, HD 149757 is known to have two strong radial velocity components at approximately  $-27$  and  $-15 \text{ km s}^{-1}$ , but the one at  $-15 \text{ km s}^{-1}$  has much larger column densities (Herbig 1968, see also Fig. 1). For the targets with only ELODIE spectra, the K I lines are outside the available wavelength range, and thus we could only use the Na I D lines to inspect the number of radial velocity components. An obvious and inherent drawback of using these lines is that they are easily saturated. We searched for Ca I and CH lines too, but in most cases, these were too weak to be seen.

With this exercise, we established that 33 of our targets can be identified as single-cloud lines of sight. However, for three of them, the UVES archival spectra have

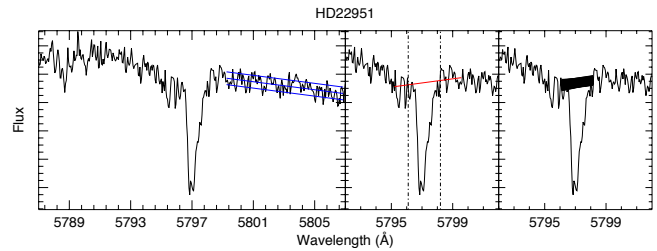


**Figure 1.** Interstellar lines for one of our targets, HD 149757 in the heliocentric rest frame. Here, the Na I lines show a dominant component at approximately  $-15 \text{ km s}^{-1}$ , as well as a second component that is much weaker near  $-27 \text{ km s}^{-1}$ . The Na I lines are somewhat saturated, but the K I lines confirm that there is just a single, dominant radial velocity component. This star meets our criterion for a single-cloud line of sight. Note that the feature near  $+27 \text{ km s}^{-1}$  in the K I (7665 Å) plot is a telluric oxygen line and not of interstellar origin.

a gap in the wavelength coverage from approximately 5760–5830 Å and thus two important DIBs,  $\lambda 5780$  and  $\lambda 5797$  are missing. We therefore excluded these targets from our data set. After these considerations, we were left with a sample of 30 single-cloud lines of sight. These targets are listed in Table 1 along with their line of sight parameters we will use in this paper (see below).

It should be noted that six of the targets in our sample – HD 23630, HD 24534, HD 110432, HD 149757, HD 164284, and HD 202904 – are Be stars. Such objects have an intrinsic  $E(B-V)$  value relative to non-Be stars of the same spectral type (Schild 1978; Sigut & Patel 2013); the hot, circumstellar gas produces excess emission in the V filter and therefore, their  $E(B-V)$  values are over-estimated. Furthermore, any dust existing in the circumstellar shell (CS) can contribute to  $E(B-V)$ , while there is no evidence to suggest that DIBs exist in CS environments (Krelowski & Sneden 1995; Snow & Wallerstein 1972; Snow 1973). Hence, we expect weaker-than-normal DIB strengths relative to  $E(B-V)$  for these six targets.

We applied a heliocentric correction to all targets, and then shifted all spectra to their interstellar rest frames. Interstellar velocity components were obtained from the literature, or measured from a known interstellar fea-



**Figure 2.** Illustration of measuring the equivalent width of the  $\lambda 5797$  DIB for the target HD 22951. (Left:) In a featureless part of the spectrum near the feature, we measured the S/N by determining the standard deviation of flux values across a straight line. The blue parallel lines show the mean flux value and the  $\pm 1\sigma$  values. (Center:) We chose a point on either side of the feature, and determined a linear continuum between the two (shown in red). The chosen integration limits are shown as dotted lines. (Right:) This figure shows the 1,000 instances of the continuum slopes generated through our Monte Carlo simulation, superimposed over the spectrum.

ture, if not available. These velocities are listed in column 14 of Table 1.

## 2.2. DIB Measurements

For our purposes, we only wanted to include those DIBs whose equivalent widths can be confidently measured, i.e., with small relative errors. This limits our selection to fairly strong and often narrow DIBs that are as much as possible free of contamination from stellar lines. The only exceptions we made was to include four of the  $C_2$  DIBs –  $\lambda\lambda 4964$ , 5513, 5546, and 5769 – despite the latter three being quite weak. We thus include a sample of 16 DIBs in our analysis:  $\lambda\lambda 4428$ , 4964, 5494, 5513, 5545, 5546, 5769, 5780, 5797, 5850, 6196, 6270, 6284, 6376, 6379, and 6614.

We measured the equivalent widths for these 16 DIBs in all lines of sight. Most of these were straightforward to measure, as they are not heavily contaminated by stellar and/or telluric features. The largest uncertainty on these measurement stems from establishing the continuum. To obtain a good estimate of these uncertainties, we used the following Monte Carlo approach to vary the continuum level and perform direct integration of the spectra (similar to that in Bhatt & Cami (2015); see Figure 2 for illustration). First, we measured the standard deviation of the flux values over a specified, featureless range of data in the vicinity of the feature to estimate the uncertainty on the flux values – i.e. we measured the signal-to-noise ratio (S/N). Next, we selected a point on each side of the feature and defined a continuum baseline by adopting a linear continuum between those two points. We also selected points as our integration limits; note that for consistency, we used the

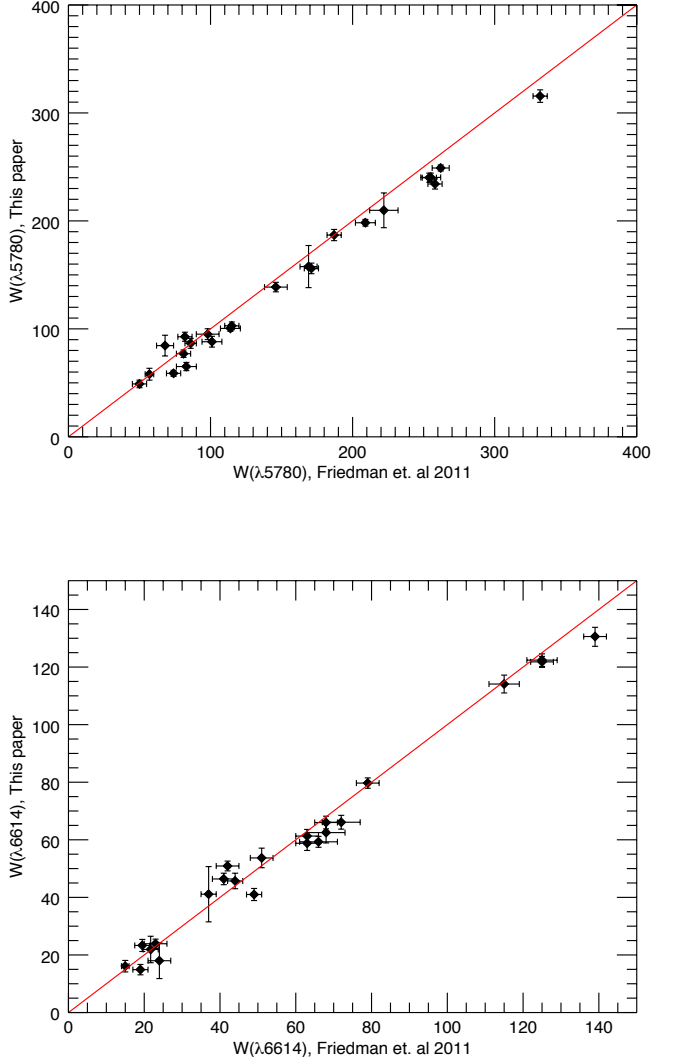
same integration limits for the same feature in all lines of sight. To simulate the process of determining the continuum, we varied the two selected continuum points 1,000 times by adding a random number to the flux values selected from a normal distribution with a mean of zero and standard deviation equal to  $\frac{1}{3}$  the measured standard deviation in the featureless continuum; we believe that this represents well how accurately one can position the continuum (which corresponds to determining the mean flux in the adjacent continuum), and we found that this produces reasonable continuum estimates (see Figure 2). Using one full standard deviation produces many continuum points which are clearly too high or too low and thus result in unrealistic continuum levels. In essence, we thus simulated the entire process of determining a continuum line 1,000 times. For each continuum, we then measured the equivalent width. The equivalent width we use in this paper is then the mean of these 1,000 measurements, and the standard deviation of these measurements provides the uncertainty.

We kept a few precautions in mind when using this method. For instance, the sharp and narrow  $\lambda 5797$  is known to be blended with the broader and shallower  $\lambda 5795$ . To avoid measuring a contribution from  $\lambda 5795$ , we measured the  $\lambda 5797$  while treating the  $\lambda 5795$  as continuum (see Figure 2). If bad pixels were found within the integration range, that data point was replaced with the average of the neighboring points. For HD 23180, the  $\lambda 5494$  feature was strongly contaminated by a stellar line. We could not resolve the two features to obtain a proper measurement, so instead we adopted the value obtained from higher resolution observations by Bondar (2012).

The two broad DIBs in our sample –  $\lambda 4428$  and  $\lambda 6284$  – cannot be measured using the methods described above, because they are heavily contaminated by stellar and telluric features, respectively. For the  $\lambda 6284$  DIB, we first applied a telluric correction using *molecfit* (version 1.1.0) (Smette et al. 2015; Kausch et al. 2015) and then proceeded as for the other DIBs.

The  $\lambda 4428$  DIB is extremely broad, and the region spanned by this DIB is plagued by stellar features. Snow et al. (2002) showed that the intrinsic profile of the band is Lorentzian, and thus, rather than numerically integrating the DIB profile, we preferred to fit a Lorentzian profile to the observations and determine the equivalent widths from the fitted parameters. In addition to our best fit, we also determined Lorentzians that represent the upper and lower envelope of the observed profiles; these then have a different full-width-at-half-max (FWHM) and central depth (CD) value (while we kept the same continuum). We found the difference between

the best-fit values and upper or lower envelope values and determined the uncertainties on EW through error propagation.



**Figure 3.** A comparison of our measured  $\lambda 5780$  and  $\lambda 6614$  equivalent width values (vertical axes) to those found in Friedman et al. (2011) (horizontal axes). The two data sets agree very well; the correlation coefficient between our  $W(\lambda 5780)$  values is  $r = 0.995$ , while that of  $W(\lambda 6614)$  is  $r = 0.993$ .

We compared our measurements to values found in the literature whenever possible and found a generally good agreement. For instance, 23 out of our 30 lines of sight were also studied by Friedman et al. (2011) and we found a very good correlation between their reported EW values and ours (see e.g. Figure 3).



All of our equivalent width measurements are shown in Table 5 in the appendix to this paper.

### 2.3. Line of Sight Parameters

To have a better chance of understanding what drives the variations in the DIB strengths, we need to include other parameters that offer some description of the lines of sight. The Jenkins (2009) study provides a critically reviewed set of measurements that describe the lines of sight we study here and we reproduce some of their values in Table 1.

There are two quantities that are related to the dust in the line of sight. The amount of dust is traditionally characterized by using the color excess,  $E(B-V)$ , and is often used as a normalization factor for DIB strengths. We adopt an uncertainty of  $\pm 0.02$  mag for our  $E(B-V)$  measurements. Jenkins (2009) furthermore showed that the depletion of different elements in a line of sight can be described by a single parameter,  $F_\star$ , that is a measure for the “total depletion”; individual elemental depletion factors scale with  $F_\star$ . Since depletion may play a role in the formation and/or destruction of the DIB carriers, we include it here in our analysis.

Information on the gas in the line of sight stems from measurements of hydrogen, in particular the column densities of neutral and molecular hydrogen,  $N(\text{H I})$  and  $N(\text{H}_2)$ , respectively. Note that total hydrogen column densities  $N(\text{H})$  are not listed, but can easily be calculated as  $N(\text{H}) = N(\text{H I}) + 2N(\text{H}_2)$ . For HD 23630, a reliable  $N(\text{H I})$  value could not be obtained by Jenkins; consequently, he uses synthetically derived  $N(\text{H})$  and  $F_\star$  values, which we adopt for this paper. We use this synthetic  $N(\text{H})$  value and the measured  $N(\text{H}_2)$  value to derive a synthetic  $N(\text{H I})$  from  $N(\text{H}) - 2N(\text{H}_2)$ . For two other targets, HD 27778 and HD 202904, Jenkins provides only upper limits and best values for  $N(\text{H I})$ . We take the lower limit to be zero in both cases. For  $N(\text{H})$ , he lists only upper and lower limits. We adopted  $N(\text{H})_{\text{best}} = N(\text{H I})_{\text{best}} + 2N(\text{H}_2)_{\text{best}}$  for these targets, as well as synthetic  $F_\star$  values. For HD 35149, only an upper limit for  $N(\text{H}_2)$  is provided. In this case,  $N(\text{H}_2)$  is quite small, so Jenkins takes  $N(\text{H}) = N(\text{H I})$  and calculates  $F_\star$  normally. Because  $N(\text{H}_2)$  is so small in this case, we take Jenkins’ value to be the best value, set the lower limit equal to zero, and set the upper limit equal to the best value. It has been suggested that  $f(\text{H}_2)$  can be used as an indicator for the amount of interstellar UV radiation that can penetrate since  $\text{H}_2$  is dissociated when not shielded (see e.g. Cami et al. 1997; Sonnentrucker et al. 1997). This could be an important parameter in our study, and we therefore calculated the fraction of

molecular hydrogen,  $f(\text{H}_2)$ , for each line of sight from  $N(\text{H}_2)$  and  $N(\text{H})$ .

Similarly, it has long been known that the ratio  $W(\lambda 5797)/W(\lambda 5780)$  is somehow related to the physical conditions in a region of space (Krelowski et al. 1997; Cami et al. 1997); more specifically, it is also suggested to trace UV exposure, where larger values of the  $W(\lambda 5797)/W(\lambda 5780)$  ratio correspond to more sheltered ( $\zeta$ -type) environments. We thus also include the  $W(\lambda 5797)/W(\lambda 5780)$  ratio as a line of sight parameter in our study.

Our sample is quite diverse in terms of  $f(\text{H}_2)$ : reported  $f(\text{H}_2)$  values in diffuse clouds run from 0 to  $\sim 0.8$  (Snow & McCall 2006); our sample covers a similar range from 0 to 0.824. Our sample also covers a broad range of physical conditions, as it includes both archetypal  $\sigma$  and  $\zeta$  sightlines. For comparison, our  $W(\lambda 5797)/W(\lambda 5780)$  values range from 0.10 to 0.65 whereas those calculated from EWs in a large sample of 133 stars by Friedman et al. (2011) range from 0.15 to 0.85. Furthermore, the boundary between  $\sigma$  and  $\zeta$  clouds is usually placed around  $W(\lambda 5797)/W(\lambda 5780) = 0.3$  and therefore, we include a large sample of both cloud types. For  $F_\star$  we similarly cover much of the possible range. Jenkins defines  $F_\star$  such that a line of sight with minimal depletions has a value of 0, and the  $-15 \text{ km s}^{-1}$  component of HD 149757 – the archetype for *large* depletions – has a value of 1.0. Our values for  $F_\star$  are found between 0.37 and 1.19. For  $E(B-V)$ , the situation is slightly more complicated. Because we restrict ourselves to single clouds, we are biased towards objects of lower reddening. Indeed, our  $E(B-V)$  values range from 0.00 to 0.47, while it is not uncommon for multiple-cloud sightlines to exceed 1. In particular,  $E(B-V)$  values below 0.08 can be problematic; below this threshold,  $\text{H}_2$  does not exist in appreciable amounts (Savage et al. 1977) and many DIBs similarly fall below the limit of detection. Within our sample there are six of these low- $E(B-V)$  targets: HD 23630, HD 24760, HD 35715, HD 36822, HD 143275, and HD 214680. Studies by Kumar et al. (1982), Federman et al. (1984), and Kumar (1986) confirm that strong DIBs such as  $\lambda 5780$ ,  $\lambda 5797$ ,  $\lambda 6284$ , and  $\lambda 6614$  can be detected toward such objects. Moreover, Galazutdinov et al. (1998) confirm that  $W(\lambda 5797)/W(\lambda 5780)$  is still sufficient to differentiate between  $\sigma$  and  $\zeta$  environments, despite the weak reddening. Some weaker DIBs may not appear in these poorly reddened objects; in such cases, sightlines having larger  $E(B-V)$  values will still be observable and will dominate the overall trends.

Table 1. Basic target data.

Target	Alt.	RA	DEC	V	E(B-V)	N(H I)	N(H <sub>2</sub> )	f(H <sub>2</sub> )	F <sub>★</sub>	$\frac{W(\lambda 5797)}{W(\lambda 5780)}$	$v_{\text{ISM}}^a$	Ref <sup>a</sup>	Data
	Name	[J2000]	[J2000]			[10 <sup>21</sup> cm <sup>-2</sup> ]	[10 <sup>20</sup> cm <sup>-2</sup> ]				[km s <sup>-1</sup> ]		Source
HD 15137		02 27 59.81	+52 32 57.6	7.86	0.24	1.29 <sup>+0.57</sup> <sub>-0.40</sub>	1.86 <sup>+0.26</sup> <sub>-0.12</sub>	0.22 <sup>+0.09</sup> <sub>-0.06</sub>	0.37±0.09	0.30±0.02	-9.58	1	ELODIE
HD 22951	40 Per	03 42 22.65	+33 57 54.1	4.98	0.19	1.10 <sup>+0.35</sup> <sub>-0.32</sub>	2.88 <sup>+1.48</sup> <sub>-0.98</sub>	0.35 <sup>+0.27</sup> <sub>-0.18</sub>	0.73±0.05	0.35±0.02	12.47	1	ELODIE
HD 23180	o Per	03 44 19.13	+32 17 17.7	3.86	0.22	0.76 <sup>+0.36</sup> <sub>-0.23</sub>	3.98 <sup>+1.64</sup> <sub>-1.16</sub>	0.51 <sup>+0.33</sup> <sub>-0.24</sub>	0.84±0.06	0.65±0.04	13.45	2	ELODIE
HD 23630	η Tau	03 47 29.08	+24 06 18.5	2.87	0.05	0.22 <sup>+0.10</sup> <sub>-0.07</sub>	0.35 <sup>+0.18</sup> <sub>-0.12</sub>	0.28 <sup>+0.23</sup> <sub>-0.15</sub>	0.89±0.10	0.16±0.04	16.76	2	ELODIE
HD 24398	ζ Per	03 54 07.92	+31 53 01.1	2.88	0.27	0.63 <sup>+0.06</sup> <sub>-0.07</sub>	4.68 <sup>+2.40</sup> <sub>-1.59</sub>	0.59 <sup>+0.46</sup> <sub>-0.31</sub>	0.88±0.05	0.55±0.02	14.54	2	ELODIE
HD 24534	X Per	03 55 23.08	+31 02 45.0	6.10	0.31	0.54 <sup>+0.08</sup> <sub>-0.07</sub>	8.32 <sup>+0.80</sup> <sub>-0.73</sub>	0.76 <sup>+0.13</sup> <sub>-0.11</sub>	0.90±0.06	0.62±0.04	14.5	5	ELODIE
HD 24760	ε Per	03 57 51.23	+40 00 36.8	2.90	0.07	0.25 <sup>+0.05</sup> <sub>-0.05</sub>	0.33 <sup>+0.27</sup> <sub>-0.15</sub>	0.21 <sup>+0.25</sup> <sub>-0.14</sub>	0.68±0.04	0.18±0.02	7.06	2	ELODIE
HD 24912	ξ Per	03 58 57.90	+35 47 27.7	4.04	0.26	1.29 <sup>+0.26</sup> <sub>-0.24</sub>	3.39 <sup>+1.40</sup> <sub>-0.99</sub>	0.35 <sup>+0.21</sup> <sub>-0.15</sub>	0.83±0.02	0.26±0.01	11.2	1	ELODIE
HD 27778	62 Tau	04 23 59.76	+24 18 03.6	6.33	0.34	0.22 <sup>+0.55</sup> <sub>-0.22</sub>	5.25 <sup>+1.06</sup> <sub>-0.88</sub>	0.82 <sup>+0.45</sup> <sub>-0.27</sub>	1.19±0.07	0.43±0.03	15.22	2	ELODIE
HD 35149	23 Ori	05 22 50.00	+03 32 40.0	5.00	0.08	0.43 <sup>+0.12</sup> <sub>-0.13</sub>	0.03 <sup>+0.00</sup> <sub>-0.03</sub>	0.02 <sup>+0.00</sup> <sub>-0.02</sub>	0.54±0.11	0.20±0.04	24.09	2	UVES
HD 35715	ψ Ori	05 26 50.23	+03 05 44.4	4.60	0.03	0.31 <sup>+0.13</sup> <sub>-0.13</sub>	6±2 × 10 <sup>-6</sup>	4±2 × 10 <sup>-6</sup>	0.66±0.11	0.10±0.04	25.2	1	ELODIE
HD 36822	φ <sup>1</sup> Ori	05 34 49.24	+09 29 22.5	4.40	0.07	0.65 <sup>+0.13</sup> <sub>-0.12</sub>	0.21 <sup>+0.09</sup> <sub>-0.06</sub>	0.06 <sup>+0.04</sup> <sub>-0.03</sub>	0.74±0.08	0.19±0.04	25.53	1	ELODIE
HD 36861	λ Ori A	05 35 08.28	+09 56 03.0	3.30	0.10	0.60 <sup>+0.16</sup> <sub>-0.16</sub>	0.13 <sup>+0.08</sup> <sub>-0.05</sub>	0.04 <sup>+0.04</sup> <sub>-0.02</sub>	0.57±0.04	0.48±0.04	25.2	3	ELODIE
HD 40111	139 Tau	05 57 59.66	+25 57 14.1	4.82	0.10	0.79 <sup>+0.16</sup> <sub>-0.15</sub>	0.54 <sup>+0.31</sup> <sub>-0.20</sub>	0.12 <sup>+0.10</sup> <sub>-0.07</sub>	0.49±0.04	0.20±0.04	15.29	2	ELODIE
HD 110432	BZ Cru	12 42 50.27	-63 03 31.0	5.32	0.39	0.71 <sup>+0.29</sup> <sub>-0.21</sub>	4.37 <sup>+0.42</sup> <sub>-0.38</sub>	0.55 <sup>+0.13</sup> <sub>-0.11</sub>	1.17±0.11	0.25±0.01	6.8	3	UVES
HD 143275	δ Sco	16 00 20.01	-22 37 18.1	2.29	0.00	1.41 <sup>+0.29</sup> <sub>-0.29</sub>	0.26 <sup>+0.15</sup> <sub>-0.09</sub>	0.03 <sup>+0.03</sup> <sub>-0.02</sub>	0.90±0.03	0.19±0.02	-10.90	2	UVES
HD 144217	β <sup>1</sup> Sco	16 05 26.23	-19 48 19.6	2.62	0.18	1.23 <sup>+0.12</sup> <sub>-0.11</sub>	0.68 <sup>+0.10</sup> <sub>-0.09</sub>	0.10 <sup>+0.02</sup> <sub>-0.02</sub>	0.81±0.02	0.11±0.01	-8.95	2	UVES
HD 145502	ν Sco	16 11 59.74	-19 27 38.5	4.13	0.20	1.17 <sup>+0.56</sup> <sub>-0.59</sub>	0.78 <sup>+0.32</sup> <sub>-0.23</sub>	0.12 <sup>+0.08</sup> <sub>-0.07</sub>	0.80±0.11	0.18±0.01	-8.49	2	ELODIE
HD 147165	σ Sco	16 21 11.32	-25 35 34.0	2.91	0.31	2.19 <sup>+0.90</sup> <sub>-0.87</sub>	0.62 <sup>+0.25</sup> <sub>-0.18</sub>	0.05 <sup>+0.04</sup> <sub>-0.03</sub>	0.76±0.06	0.13±0.01	-6.26	2	UVES
HD 147933	ρ Oph A	16 25 35.10	-23 26 48.7	5.02	0.37	4.27 <sup>+1.53</sup> <sub>-0.80</sub>	3.72 <sup>+1.53</sup> <sub>-1.09</sub>	0.15 <sup>+0.09</sup> <sub>-0.07</sub>	1.09±0.08	0.27±0.03	-8.02	2	UVES
HD 149757	ζ Oph	16 37 09.54	-10 34 01.5	2.58	0.29	0.52 <sup>+0.02</sup> <sub>-0.04</sub>	4.47 <sup>+0.90</sup> <sub>-0.75</sub>	0.63 <sup>+0.20</sup> <sub>-0.17</sub>	1.05±0.02	0.50±0.04	-14.98	2	UVES
HD 164284	66 Oph	18 00 15.80	+04 22 07.0	4.78	0.11	0.42 <sup>+0.23</sup> <sub>-0.39</sub>	0.71 <sup>+0.29</sup> <sub>-0.21</sub>	0.25 <sup>+0.18</sup> <sub>-0.20</sub>	0.89±0.18	0.15±0.02	-15.32	1	ELODIE
HD 170740		18 31 25.69	-10 47 45.0	5.76	0.38	1.07 <sup>+0.59</sup> <sub>-0.47</sub>	7.24 <sup>+1.47</sup> <sub>-1.22</sub>	0.58 <sup>+0.22</sup> <sub>-0.18</sub>	1.02±0.11	0.26±0.01	-12.9	6	UVES
HD 198478	55 Cyg	20 48 56.29	+46 06 50.9	4.86	0.43	2.04 <sup>+0.84</sup> <sub>-0.63</sub>	7.41 <sup>+3.06</sup> <sub>-2.17</sub>	0.42 <sup>+0.27</sup> <sub>-0.20</sub>	0.81±0.05	0.24±0.01	-10.04	2	ELODIE
HD 202904	ν Cyg	21 17 55.08	+34 53 48.8	4.43	0.09	0.23 <sup>+0.21</sup> <sub>-0.23</sub>	0.14 <sup>+0.07</sup> <sub>-0.05</sub>	0.11 <sup>+0.12</sup> <sub>-0.10</sub>	0.39±0.11	0.13±0.05	-12.90	4	ELODIE
HD 207198		21 44 53.28	+62 27 38.0	5.96	0.47	3.39 <sup>+0.59</sup> <sub>-0.50</sub>	6.76 <sup>+0.65</sup> <sub>-0.59</sub>	0.28 <sup>+0.05</sup> <sub>-0.05</sub>	0.90±0.03	0.53±0.01	-15.28	2	ELODIE
HD 209975	19 Cep	22 05 08.79	+62 16 47.3	5.11	0.27	1.29 <sup>+0.41</sup> <sub>-0.38</sub>	1.20 <sup>+0.62</sup> <sub>-0.41</sub>	0.16 <sup>+0.12</sup> <sub>-0.09</sub>	0.57±0.26	0.31±0.01	-11.39	2	ELODIE
HD 214680	10 Lac	22 39 15.68	+39 03 01.0	4.88	0.08	0.50 <sup>+0.14</sup> <sub>-0.15</sub>	0.17 <sup>+0.05</sup> <sub>-0.04</sub>	0.06 <sup>+0.03</sup> <sub>-0.03</sub>	0.50±0.06	0.34±0.02	-9.2	1	ELODIE
HD 214993	12 Lac	22 41 28.65	+40 13 31.6	5.23	0.06	0.58 <sup>+0.20</sup> <sub>-0.18</sub>	0.43 <sup>+0.22</sup> <sub>-0.14</sub>	0.13 <sup>+0.10</sup> <sub>-0.07</sub>	0.68±0.10	0.17±0.02	-9.44	1	ELODIE
HD 218376	1 Cas	23 06 36.82	+59 25 11.1	4.84	0.16	0.89 <sup>+0.28</sup> <sub>-0.26</sub>	1.41 <sup>+0.73</sup> <sub>-0.48</sub>	0.24 <sup>+0.19</sup> <sub>-0.13</sub>	0.60±0.06	0.28±0.01	-12.65	1	ELODIE

NOTE—RA and DEC are taken from SIMBAD.  $V$ ,  $E(B-V)$ ,  $N(\text{H I})$ ,  $N(\text{H}_2)$  and  $F_\star$  values are from Jenkins (2009) where we assume an uncertainty of  $\pm 0.02$  mag on the  $E(B-V)$  values.  $f(\text{H}_2) = 2N(\text{H}_2)/[N(\text{H}) + 2N(\text{H}_2)]$  is the fraction of molecular hydrogen.

### 3. PRINCIPAL COMPONENT ANALYSIS

Several parameters are known to contribute to DIB strength; however, most of these parameters are interrelated (e.g., a larger E(B-V) implies larger column densities of all species). For this work, we want to extract a set of uncorrelated parameters that best describe observed DIB strengths. To achieve this goal, we perform a principal component analysis (PCA).

PCA is a statistical technique used to analyze high-dimensional data. The objectives of PCA are two-fold: (1) to reduce the number of dimensions in a data set, and (2) to identify hidden patterns in a set of data. We will primarily focus on the former in this analysis. For derivations of PCA see [Pearson \(1901\)](#) or [Hotelling \(1933\)](#); for modern reviews, see [Abdi & Williams \(2010\)](#) or [Jolliffe \(2014\)](#); for uses of PCA in astronomy see e.g. [Yip et al. \(2004\)](#); [Suzuki \(2006\)](#); [Conselice \(2006\)](#); [Budavári et al. \(2009\)](#); [Páris et al. \(2011\)](#).

In this Section, we first define some of the terminology and notation associated with PCA and provide a basic mathematical overview of the process. We then discuss some of the problems and underlying assumptions associated with PCA, and how we address each of them. Finally, we provide a simple example using two variables to better illustrate our methodology and facilitate the interpretation of our results.

#### 3.1. Definitions and Terminology

The starting point of our analysis is a set of  $n$  variables for which we have  $m$  measurements  $x_i$ ; in our case, we use  $n = 23$  variables (representing DIB equivalent widths and line of sight parameters) that are measured for  $m = 30$  lines of sight. Our measurements thus span an  $n$ -dimensional parameter space – each dimension corresponding to a different variable – and a full set of measurements for a single line of sight can be represented by an  $n \times 1$  vector in this parameter space. Starting from this set of  $n$  possibly correlated (and thus not necessarily orthogonal) variables, a PCA finds a coordinate transformation that casts the variables in terms of a new, *orthogonal*,  $n$ -dimensional reference frame:

$$y_i \hat{\mathbf{n}}'_i = a_{i,1}x_1\hat{\mathbf{n}}_1 + a_{i,2}x_2\hat{\mathbf{n}}_2 + \dots + a_{i,n}x_n\hat{\mathbf{n}}_n \quad (1)$$

where the  $\hat{\mathbf{n}}_i$  represent unit vectors in the original parameter space and  $\hat{\mathbf{n}}'_i$  unit vectors in the new reference frame – these are the so-called principal components (PCs). The coefficients,  $a_{i,j}$ , are constrained such that their squared sums equal one:

$$a_{i,1}^2 + a_{i,2}^2 + \dots + a_{i,n}^2 = 1 \quad (2)$$

The coordinate transformation is chosen such that when representing the measurements in the new refer-

ence frame, the largest amount of variance in the data occurs along the axis defined by the first PC, and the second PC accounts for as much of the remaining variation in the data as possible, with the constraint that it is orthogonal to PC<sub>1</sub>. Each successive PC has the same properties: they are linear combinations of the original variables, and each accounts for as much of the remaining variation as possible, while being orthogonal to (and thus, uncorrelated with) all previous PCs.

The entire coordinate transformation can be written in matrix notation:

$$Y = AX \quad (3)$$

where  $X$  is an  $n \times m$  matrix containing the original set of data;  $A$  is the  $n \times n$  transformation matrix; and  $Y$  is an  $n \times m$  matrix containing the transformed data points in the new reference frame defined by the set of PCs<sup>1</sup>.

The PCA obtains the transformation matrix  $A$  from the eigenvalues ( $\lambda_1, \lambda_2, \dots, \lambda_n$ ) of the covariance matrix of the original variables. The rows of this matrix  $A$  are the corresponding eigenvectors; each eigenvector describing the projections of the original variables onto the new coordinate system defined by the PCs. The eigenvalues are furthermore ordered such that  $\lambda_1$  is the largest and  $\lambda_n$  is the smallest with  $\sum_i^n \lambda_i = n$ . Each eigenvalue indicates how many original variables a single PC can account for, i.e., an eigenvalue of 2 indicates that the corresponding PC carries the same weight as two of the original variables.

One of the key features of a PCA is the ability to reduce the number of dimensions in a data set. Indeed, it is often the case that we can use the results of a PCA to accurately describe the variation in a data set with a reduced number of variables. As such, PCs that account for very little variation are often ignored, and only the leading  $p$  components are kept. Selecting a value for  $p$  is somewhat arbitrary, but the decision is usually based on one of three criteria: (1) once a certain amount of variation is accounted for (e.g., 90%), all further PCs are ignored; (2) only PCs with eigenvalues greater than one are kept (representing variables that can replace more than one of the original variables); or, (3) if an “elbow” (a sharp, sudden drop-off) is noticed in a so-called “screeplot” (i.e., a plot of eigenvalue vs. component number, see e.g. [Fig. 6](#)), all PCs beyond the elbow are ignored ([Jolliffe 2014](#)). If  $p$  components are kept, then the last  $n - p$  rows of  $A$  are eliminated, and  $Y$  becomes a  $p \times m$  matrix. At that point, we effectively express the original  $n$  variables with only  $p$  new vari-

<sup>1</sup> Note that in the PCA analysis by [Suzuki \(2006\)](#), their Eq. 1 corresponds to the inverse of our Eq. 3.



ables, while keeping as much of the variance in the data as desired.

### 3.2. Treatment of Data

The variables  $x_i$  that we want to include in our analysis are listed in column 1 of Table 2. Since PCA is concerned with linear combinations of variables, it is redundant to include  $N(H)$  in addition to  $N(H\ I)$  and  $N(H_2)$ ; however, whereas  $N(H\ I)$  and  $N(H_2)$  measure the column densities of individual species,  $N(H)$  approximates the total amount of gas in the line of sight. Recognizing this distinction, we choose to keep  $N(H)$  in addition to  $N(H\ I)$  and  $N(H_2)$ .

There are a few underlying assumptions associated with PCA, which we will discuss. The first is that raw data is comparable in units and magnitude; otherwise, the results of PCA will be more strongly influenced by variables that have larger variances (i.e., variations among  $N(H)$  values which are on the order of  $10^{21}$  will completely dominate those of  $E(B-V)$ , which are on the order of  $10^{-1}$ ). To address this problem, we standardize each variable prior to performing PCA:

$$z_{i,j} = \frac{x_{i,j} - \bar{x}_i}{s_{x_i}} \quad (4)$$

where  $i$  refers to the  $i^{th}$  variable and  $j$  refers to the  $j^{th}$  observation. Essentially, we subtract the mean and divide the residuals by the standard deviation, and use the resulting variables  $z_{i,j}$  as input to the actual PCA. Columns 2 and 3 of Table 2 show the mean and standard deviation of each of our variables, respectively. After this process, the standardized variables ( $z_1, z_2, \dots, z_{23}$ ) each have a mean of zero and a standard deviation of one.

A second assumption of PCA is that all values are accurately known. This assumption is challenged by the fact that our measurements have uncertainties. Since PCA does not consider uncertainties, we need a way to quantify how reliable our results are. Since the PCs represent vectors in a multi-dimensional parameter space, this is not straightforward.

We therefore used a Monte Carlo (MC) simulation to estimate the uncertainties on our results. We first performed PCA using our measured values. These results provide the PCs that we discuss in the sections that follow. Next, we generated 1,000 perturbed data sets, by adding random noise to each observation, selected from a normal distribution with a mean of zero and a standard deviation corresponding to the error bar on each measurement. By definition, none of our measured quantities can be negative, so if a perturbed measurement was negative after the addition of random noise,

the value was replaced with zero. For each of these new data sets, we standardized the variables<sup>2</sup> and performed a separate PCA. The result is 1,000 unique transformation matrices. For each entry in the transformation matrix, we found the mean ( $\bar{a}_{i,j}$ ) and standard deviation ( $s_{a_{i,j}}$ ) among the 1,000 unique matrices. Then we computed the upper and lower limit on each value in the matrix  $A$  according to  $\bar{a}_{i,j} \pm s_{a_{i,j}}$ . We compared these upper and lower limits to the original, unperturbed results and ultimately, were left with positive and negative errors on each entry in the transformation matrix (i.e., each component of each eigenvector). We applied the same methods to obtain uncertainty measurements on each entry in  $Y$ , the matrix of transformed data points.

A third assumption associated with PCA is that each variable follows a normal distribution. We tested our variables for normality and found that in some cases, the distributions were not normal. The assumption of normality, however, is not critical and the results of PCA should still be valid if this condition is not met, so we decided to continue with the analysis.

Finally, the results of PCA can be sensitive to outliers. Despite including some unusual targets (e.g., HD 147933 is found within a dark cloud, and HD 36822 and HD 36861 are found near H II regions), we do not remove any targets from our sample. In order to probe a wide range of interstellar conditions, it is important to include these abnormal lines of sight: Any conclusions made about the DIBs through this analysis should extend to these sightlines as well. The only potential complication is for the six Be stars discussed in Section 2.1. The atypical  $E(B-V)$  values arise from circumstellar material and are therefore not a part of the interstellar DIB environment. We acknowledge that they could influence our results and therefore we differentiate these points in the plots and analyses that follow.

### 3.3. A Simple 2D Example

It is insightful to illustrate our methodology by performing a PCA on a simple, two-dimensional dataset; we therefore performed a PCA analysis using only the measurements for the variables  $E(B-V)$  and  $N(H)$ . Using the notation provided in Table 2, we denote these variables  $x_{17}$  and  $x_{20}$  and their corresponding standardized

<sup>2</sup> Note that we recalculated the means and standard deviations for each perturbed data set before standardizing the variables.

**Table 2.** Input variables used in the PCA, and their mean values and standard deviations.

Variable Name ( $x_i$ )	Mean ( $\bar{x}_i$ )	Standard Deviation ( $s_{x_i}$ )
$x_1 = W(\lambda 4428)$	645.9	350.9
$x_2 = W(\lambda 4964)$	6.8	5.7
$x_3 = W(\lambda 5494)$	5.6	4.1
$x_4 = W(\lambda 5513)$	3.9	3.9
$x_5 = W(\lambda 5545)$	5.9	4.3
$x_6 = W(\lambda 5546)$	2.8	2.2
$x_7 = W(\lambda 5769)$	2.6	2.2
$x_8 = W(\lambda 5780)$	131.5	77.0
$x_9 = W(\lambda 5797)$	37.7	27.8
$x_{10} = W(\lambda 5850)$	15.6	13.6
$x_{11} = W(\lambda 6196)$	13.5	8.3
$x_{12} = W(\lambda 6270)$	20.7	13.9
$x_{13} = W(\lambda 6284)$	149.4	84.1
$x_{14} = W(\lambda 6376)$	9.3	7.4
$x_{15} = W(\lambda 6379)$	24.7	18.0
$x_{16} = W(\lambda 6614)$	52.5	35.2
$x_{17} = E(B-V)$	0.20	0.13
$x_{18} = N(H\ I)$	$1.0 \times 10^{21}$	$9.2 \times 10^{20}$
$x_{19} = N(H_2)$	$2.4 \times 10^{21}$	$2.6 \times 10^{20}$
$x_{20} = N(H)$	$1.5 \times 10^{21}$	$1.2 \times 10^{21}$
$x_{21} = f(H_2)$	0.27	0.23
$x_{22} = F_\star$	0.75	0.21
$x_{23} = \frac{W(\lambda 5797)}{W(\lambda 5780)}$	0.2907	0.1567

NOTE—Entries for  $i=1,2,\dots,16$  refer to equivalent widths of named features in mÅ. Units for the remaining entries are those specified in Table 1. Standardized variables  $z_i$  correspond to the  $x_i$ , e.g.  $z_8$  is the standardized equivalent width of the  $\lambda 5780$  DIB.

forms as  $z_{17}$  and  $z_{20}$ . The  $2 \times 30$  matrix of (standardized) measurements  $X$  from Equation 3 is then:

$$X = \begin{bmatrix} z_{17} \\ z_{20} \end{bmatrix} = \begin{bmatrix} 0.273 & -0.106 & \dots & -0.333 \\ 0.121 & 0.121 & \dots & -0.285 \end{bmatrix} \quad (5)$$

**Table 3.** Principal components, eigenvalues and relative importance of each PC for a 2D example involving E(B-V) and N(H).

PC	Eigen- value	% Variation	Cumulative %	Eigenvector
1	1.813	90.63	90.63	(0.707, 0.707)
2	0.187	9.37	100.00	(0.707, -0.707)

From this, the PCA results in the following  $2 \times 2$  transformation matrix  $A$ :

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix} \quad (6)$$

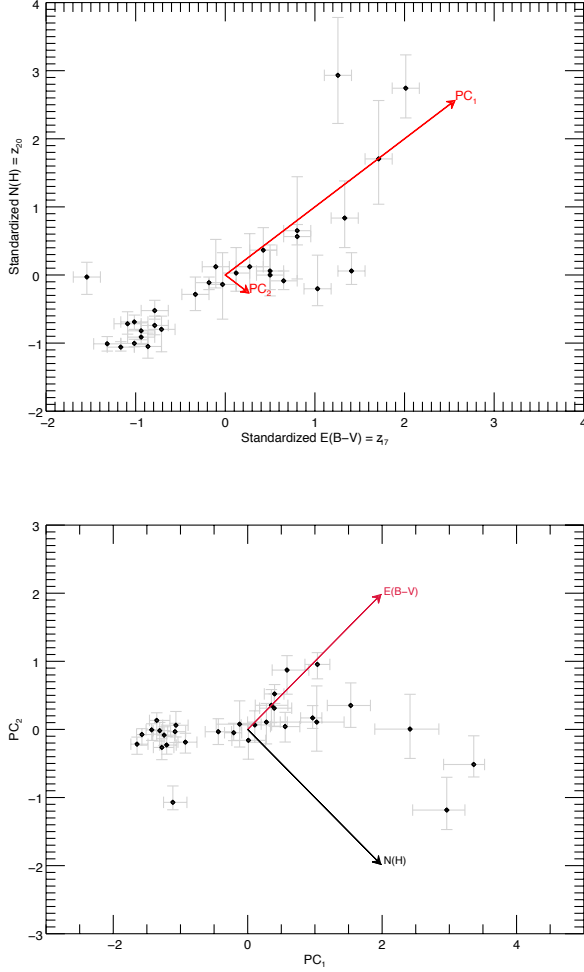
as well as the  $2 \times 30$  matrix,  $Y$ , containing the transformed data points:

$$Y = \begin{bmatrix} PC_1 \\ PC_2 \end{bmatrix} = \begin{bmatrix} 0.279 & 0.011 & \dots & -0.437 \\ 0.107 & -0.161 & \dots & -0.034 \end{bmatrix} \quad (7)$$

Note that we use the notation  $PC_i$  to describe the measurement values in the new coordinate frame described by the PCs. That is,  $PC_i = y_i$  in Eq. (1). The results pertaining to the PCs for this example are summarized in table 3. Column 1 lists the PC number, while Column 2 shows the eigenvalue associated with that PC. Since there are two original variables, the eigenvalues sum to 2. The fraction of the total variation in the data for which each PC is responsible is shown in Column 3, followed by the cumulative fraction in Column 4. The (unit length) eigenvectors are shown in Column 5.

Figure 4 illustrates these results in two ways. In the top panel, we show the original data set along with the PC vectors (i.e., the eigenvectors corresponding to  $PC_1$  and  $PC_2$ ). From this figure, it is clear that the vector  $PC_1$  lies in the direction of maximum variance. From table 3, we see that in fact almost all of the variation in the data set ( $>90\%$ ) is accounted for by  $PC_1$ . This indicates that a single variable can adequately describe variations among E(B-V) and N(H) – an interstellar finding that is well known (e.g. Bohlin et al. 1978, see also discussion below).

In the bottom panel of Figure 4, we illustrate the results in the form of a so-called “biplot”. Here, we show the transformed data points (i.e., those of matrix  $Y$ ) and the projections of the original variables onto the  $PC_1$ - $PC_2$  plane. This format is much more convenient for



**Figure 4.** (*Top:*) The original standardized data points and their uncertainties (black) and the PCs (red), the latter scaled up by a constant factor for clarity.  $PC_1$  points in the direction of maximum variation.  $PC_2$  is perpendicular to  $PC_1$  and accounts for the remaining variation.

(*Bottom:*) A so-called “biplot” illustrating the relationship between the original variables and the new set of PCs. The figure shows the transformed data points (Y) in the  $PC_1$ - $PC_2$  plane. The original variables are shown as vectors projected onto the  $PC_1$ - $PC_2$  plane. For example,  $E(B-V)$  has a projection on  $PC_1$  equal to  $a_{1,1} = 0.707$  and projection on  $PC_2$  equal to  $a_{1,2} = 0.707$ . Similarly,  $N(H)$  has a projection on  $PC_1$  equal to  $a_{2,1} = 0.707$  and a projection on  $PC_2$  equal to  $a_{2,2} = -0.707$ . Note that the vectors have been scaled for clarity in this figure too.

interpreting the PCs. For instance, we can examine the projections of  $E(B-V)$  and  $N(H)$  onto  $PC_1$  (i.e., only the horizontal components of the vectors) to see how  $PC_1$  influences each original variable. Both projections onto  $PC_1$  are positive; thus, an increase of  $PC_1$  implies an

increase of both  $E(B-V)$  and  $N(H)$ . An obvious interpretation for  $PC_1$  is therefore the amount of material in the line of sight: more material will result in larger  $E(B-V)$  and  $N(H)$  values at the same time.

Doing the same for  $PC_2$ , we see that the vertical component of the  $E(B-V)$  vector is positive, while that of  $N(H)$  is negative.  $PC_2$  therefore represents some difference between  $E(B-V)$  and  $N(H)$ . A possible interpretation is therefore the dust-to-gas ratio: an increase in the dust-to-gas ratio can result in an increase in the  $E(B-V)$  or a decrease in the  $N(H)$ .

Using the eigenvectors, we can construct equations for  $PC_1$  and  $PC_2$ . Since the PCs were constructed from the standardized variables, the resulting equations will be in terms of  $z_{17}$  and  $z_{20}$ ; using Eq. 4 and the mean and standard deviations in Table 2, however, we can express them in terms of the original variables,  $E(B-V)$  and  $N(H)$ .

$$PC_1 = 0.707(z_{17}) + 0.707(z_{20}) \quad (8)$$

$$= 5.35(E(B-V)) + (5.93 \times 10^{-22})N(H) - 1.99$$

$$PC_2 = 0.707(z_{17}) - 0.707(z_{20}) \quad (9)$$

$$= 5.35(E(B-V)) - (5.93 \times 10^{-22})N(H) - 0.19$$

We can also use our example to demonstrate how to reconstruct the observed data with reduced dimensionality. In our 2D example,  $PC_1$  accounts for most of the variation in the data, and one could argue that the variations in  $PC_2$  are mostly insignificant. In other words, we can attribute variations in  $PC_2$  to noise in the data, and thus we can describe our data by just using 1 variable:  $PC_1$ . Mathematically, we thus eliminate the final column of matrix A, forming a  $1 \times 30$  matrix, Y; in doing so, we leave Equation 8 as is and set Equation 9 equal to zero. Solving for  $N(H)$  then reveals:

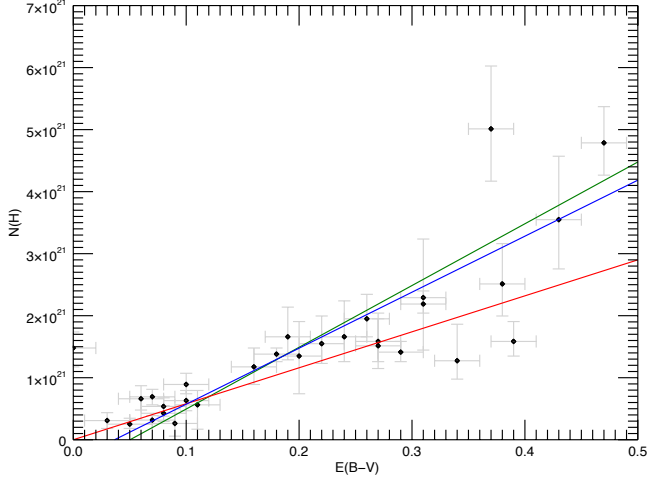
$$N(H) = 9.0 \times 10^{21} E(B-V) - 0.3 \times 10^{21} \quad (10)$$

while a least-squares fit to the original data set yields

$$N(H) = 10.0 \times 10^{21} E(B-V) - 0.50 \times 10^{21} \quad (11)$$

Both results are comparable (see the blue and green line in Fig. 5), illustrating that a PCA can be used to reliably reconstruct relationships between variables that are present in the original data set. Note though that this is somewhat different from the well-known finding by Bohlin et al. (1978):

$$N(H) = 5.8 \times 10^{21} E(B-V) \quad (12)$$



**Figure 5.**  $N(\text{H})$  (in  $\text{cm}^{-2}$ ) as a function of  $E(\text{B}-\text{V})$ . Equation 10, obtained by setting  $\text{PC}_2=0$  is represented by the blue line. The best-fit line obtained through a linear least squares fit (Equation 11) is shown in green. The well-known relation found by Bohlin et al. (1978, Eq. 12) is shown in red for comparison.

This relation is also shown in Fig. 5 (red line). Two important factors contribute to the difference between Bohlin et al. (1978) and our results. First, our sample of sightlines is biased to contain only single cloud lines of sight, with diverse physical conditions whereas Bohlin et al. (1978) used a large sample of much more reddened lines. Second, the Bohlin et al. (1978) relation was forced to go through the origin; the PCA and least squares fit were not. A discussion of such differences is not relevant for our work though.

This example shows the steps involved in carrying out a PCA and using its results. For what follows, we will be using multi-dimensional data though; the only complication is then that for the discussion and interpretation, we need to work with projections onto a 2D space for proper visualization.

#### 4. PCA RESULTS

In this section, we perform PCA using the full set of 23 variables listed in Table 2. The resulting PCs are listed in Table 4, along with their corresponding eigenvalues and the percent of the variation in the full data set for which each PC is responsible. Decidedly fewer than 23 components are required to describe the observed variations; in fact,  $\text{PC}_{23}$  is completely unnecessary, indicating that there is a perfect correlation between some of our variables (i.e., that between  $N(\text{H})$ ,  $N(\text{H I})$ , and  $N(\text{H}_2)$ ). Most of the variation is captured by the first few PCs;

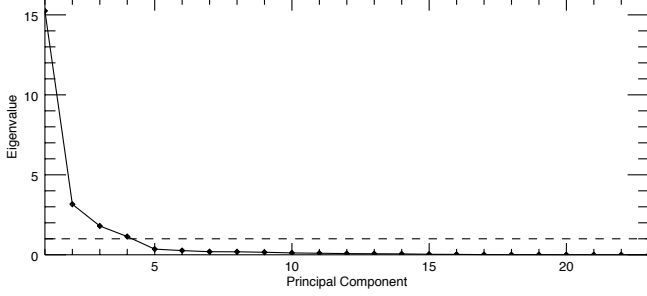
**Table 4.** Principal components, eigenvalues and relative importance of each PC.

PC	Eigenvalue	% Variation	Cumulative %
1	15.248	66.30	66.30
2	3.158	13.73	80.03
3	1.801	7.83	87.86
4	1.139	4.95	92.81
5	0.355	1.54	94.35
6	0.262	1.14	95.49
7	0.192	0.84	96.33
8	0.186	0.81	97.14
9	0.157	0.68	97.82
10	0.117	0.51	98.33
11	0.096	0.42	98.75
12	0.074	0.32	99.07
13	0.066	0.29	99.35
14	0.055	0.24	99.60
15	0.032	0.14	99.74
16	0.025	0.11	99.85
17	0.012	0.05	99.90
18	0.008	0.03	99.93
19	0.006	0.03	99.96
20	0.005	0.02	99.98
21	0.003	0.01	99.99
22	0.002	0.01	100.00
23	0.000	0.00	100.00

$\text{PC}_1$  alone accounts for over 65%, while  $\text{PC}_1$  and  $\text{PC}_2$  collectively account for over 80% of the overall variation.

To determine  $p$ , the number of significant PCs, we use criteria 2 and 3 from Section 3.1. A screeplot is shown in Figure 6, where the dashed line illustrates an eigenvalue of one. Only the first four PCs lie above this limit. For the subsequent PCs, the uncertainties obtained through our MC simulation become too large to be interpreted. This could indicate that there are no clear relations beyond this point (i.e., the variation is guided by MC perturbations rather than systematic trends in the data) or could be the result of compounding uncertainties (i.e., since PCs are constrained to be orthogonal, each PC has larger uncertainties than all previous PCs).

Rather than listing the individual components of each PC, we present our results in the form of biplots (e.g.,



**Figure 6.** A screeplot illustrating the relative importance of each PC. The dashed line indicates an eigenvalue of one. The first four PCs lie above this limit.

see Figure 7), although we include the full equations for the first four PCs in the Appendix (see Equations A1 through A4). We use our MC simulation described in Section 3.2 to draw error bars on these data. The original set of axes (i.e., the original variables) are also illustrated as vectors projected onto the two-dimensional PC plane. These are obtained through the transformation matrix,  $A$  (i.e., column  $i$  of matrix  $A$  contains the projections of the  $n$  original variables onto  $PC_i$ ). Our MC approach also provides uncertainties on these values, which we illustrate as rectangular outlines surrounding each vector head.

Below, we take a close look at the results of our PCA, and try to interpret the PCs in terms of the physical parameters that drive the variations in the DIBs. However, while each PC represents a single measurable parameter that drives part of the variations, this parameter can be related to more than one *physical* quantity; if so, these physical quantities must then necessarily be correlated. There is thus some risk in trying to interpreting the PCs in terms of single quantities. To facilitate our interpretation, we have furthermore performed two additional PCAs – one using only DIBs ( $z_1, z_2, \dots, z_{16}$ ) and using only line of sight parameters ( $z_{17}, z_{18}, \dots, z_{23}$ ); we compare the results to those obtained by including the full set of 23 variables. Each PC will be discussed separately in the sections that follow.

#### 4.1. $PC_1$ : DIB column density

Biplot for  $PC_1$  and  $PC_2$  are shown in Figure 7; the top panel shows the full PCA results while the other panels show the results of the additional PCA's using only line of sight parameters and using only DIBs. As was the case for the 2D example (see Sect. 3.3), the biplot shows the measurements for all lines of sight transformed into the PC reference frame, and only shows the values for the first two PCs – i.e. the two parameters

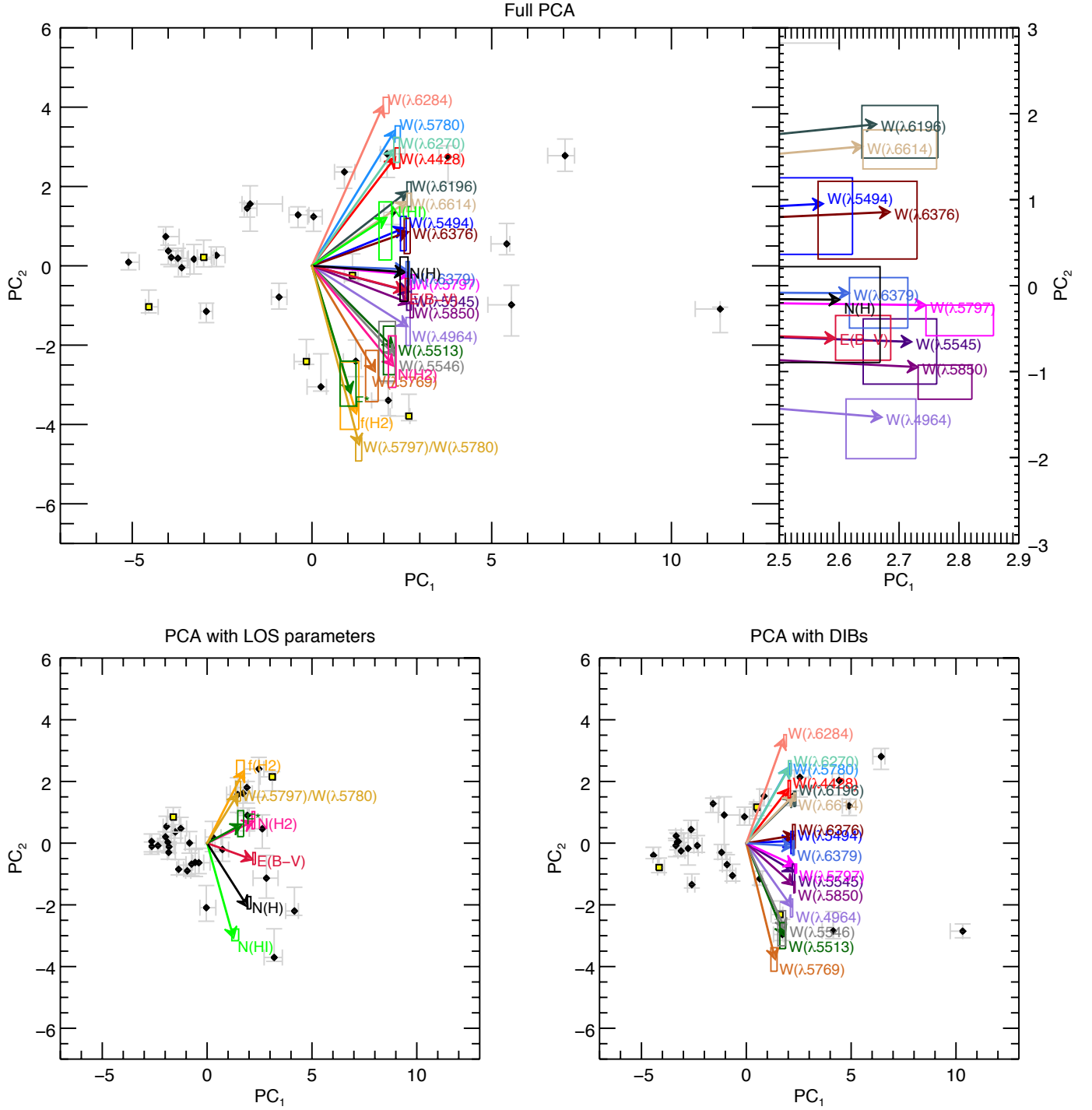
that together carry 80% of the variance in the data set. The biplot also shows the projection of the original variables onto the plane defined by these two PCs. Such projections are useful to find trends and correlations: the larger the projection of an original variable is onto an axis, the better the original variable correlates with the PC. Furthermore, original variables whose projections onto the PC plane are similar, must necessarily represent tight correlations as well. Note as well the rectangular areas at the vector heads (best seen in the close-up in the top panel to the right) that represent the uncertainties on the projections as derived from a MC simulation.

Since  $PC_1$  and  $PC_2$  capture most of the variation in the data, the  $PC_1$ - $PC_2$  biplot is particularly insightful for identifying trends and correlations. For example,  $W(\lambda 6614)$  and  $W(\lambda 6196)$  align almost perfectly with one another within their respective uncertainty ranges in  $PC_1$ . This implies that the two DIBs must correlate very tightly – a well-known fact for these DIBs (see e.g. McCall et al. 2010). Similarly,  $W(\lambda 5780)$  is known to correlate well with  $W(\lambda 6284)$ ,  $W(\lambda 6270)$ , and  $W(\lambda 4428)$  (Herbig 1993; Krelowski & Walker 1987) and these appear grouped together on the  $PC_1$ - $PC_2$  plane. Many more known relations between DIB strengths (see e.g. Lan et al. 2015, for an overview) can be recovered from these plots.

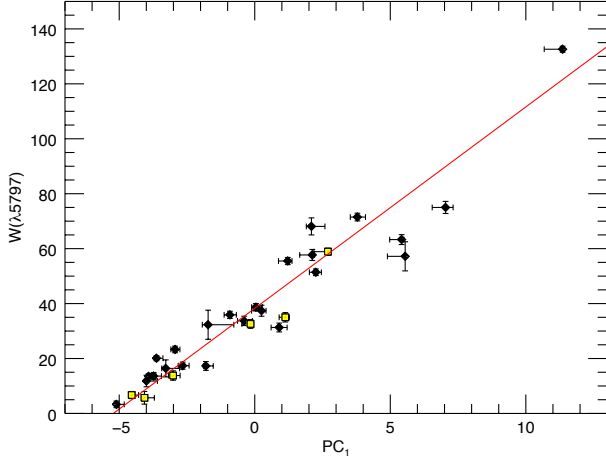
To infer clues about what  $PC_1$  physically represents, we can look at the projections of the original variables onto the  $PC_1$  axis. These projections can be easily determined from Eqs. (A1) and (A2). For example, the vector that corresponds to the projection of the  $\lambda 5780$  DIB (i.e.  $z_8$ , see Table 2) onto the  $PC_1$ - $PC_2$  plane is given by the coefficients in front of  $z_8$  in Eqs. (A1) and (A2), i.e. the corresponding vector is (0.207, 0.306). This vector is shown (albeit scaled up for visualization) in Fig. 7. As  $PC_1$  increases, so too does each of the original variables. Therefore, we interpret  $PC_1$  as measuring the amount of material in the line of sight; as the amount of material increases, so does the strength of each DIB along with variables such as color excess and the various column densities. Intuitively, this makes sense: The factor that most strongly contributes to the column density of any species in a given line of sight is the amount of material.

We can elaborate further by examining the other two PCA cases. The bottom right panel of Figure 7 shows the PCA results when only DIBs are considered. Again, each variable increases in the same direction, so  $PC_1$  must once again be some proxy for the amount of material in the line of sight. However, since we only consider DIBs in this case,  $PC_1$  is specifically tracing the amount of material that contributes to DIBs. If we com-





**Figure 7.** PC<sub>1</sub>-PC<sub>2</sub> biplots for (*Top*;) all variables, (*Bottom left*;) line of sight parameters only (excluding DIBs), and (*Bottom right*;) DIBs only, excluding line of sight parameters. The 23-dimensional vectors corresponding to the original variables are projected onto the PC<sub>1</sub>-PC<sub>2</sub> plane. The rectangular outlines surrounding the vectors indicate the uncertainty range for each projection, obtained through a MC simulation. Note that the vectors have been scaled by a constant factor for better visualization (as was done in Fig 4). Be stars are indicated by yellow squares. To help with clarity, a zoomed-in and rescaled portion of the full PCA results is presented next to the main figure. The same color scheme is used for all figures.



**Figure 8.**  $W(\lambda 5797)$  and  $PC_1$  have a very strong correlation ( $r=0.957$ ). The equation of the best-fit line is  $W(\lambda 5797) = 7.31(PC_1) + 38.25$ . Be stars are indicated by yellow squares.

pare these results to those in the top panel, we find the same ordering of DIBs – i.e.,  $W(\lambda 5797)$  still has the strongest  $PC_1$  projection and  $W(\lambda 5769)$  still has the weakest. This confirms that  $PC_1$  traces the amount of DIB-producing material in the top panel, as well. This conclusion is further established by the bottom left panel, illustrating the PCA results when only the line of sight parameters are considered. Here,  $PC_1$  appears to be quite different: Both  $N(H_2)$  and  $E(B-V)$  have stronger  $PC_1$  projections than  $N(H)$ , whereas in the top panel, the  $N(H)$  and  $E(B-V)$  projections on  $PC_1$  are comparable, while that of  $N(H_2)$  is much shorter. This means that  $PC_1$  is tracing something slightly different in each case. In the former, it seems to trace something more closely related to the amount of dust (given the large  $E(B-V)$  projection), whereas in the latter, the amount of gas plays a stronger role. Since the DIB carriers are widely accepted as begin gaseous species, it seems reasonable that  $N(H)$  would more strongly influence the strength of DIBs than  $E(B-V)$  as illustrated in our full PCA results, so this too is consistent with our interpretation that  $PC_1$  traces the amount of DIB-producing material. The fact that these cases are so different while the just-DIB results are nearly identical to those obtained using the full data set implies that our results are driven by the DIBs, rather than the line of sight parameters.

Interestingly, the largest projection onto  $PC_1$  comes from the  $\lambda 5797$  DIB, meaning that  $PC_1$  has the strongest correlation with  $W(\lambda 5797)$  ( $r = 0.957$  – see Figure 8). This result also carries over to the just-DIB case ( $r=0.963$  – not shown). Hence, the best measure

for the amount of DIB-producing material in the line of sight is  $W(\lambda 5797)$  and not  $N(H)$  or  $E(B-V)$ . Therefore,  $W(\lambda 5797)$  is conceivably a more appropriate normalization factor for DIB strengths than  $E(B-V)$ . This is discussed more in Section 5.

After establishing that  $PC_1$  traces the amount of DIB-producing material along the line of sight, we can define a new parameter,  $N_{DIB}$ , describing this value. The transformed data points derived from PCA are, by definition, centered at zero and therefore approximately half of our sightlines have negative  $PC_1$  values. For  $N_{DIB}$  to be physically meaningful, we need to have only positive values. Since our PCs were calculated from standardized variables, we have a way of adjusting our values. Using equation 4, we subtracted the mean divided by the standard deviation from each variable prior to performing PCA. To correct for this, we must add these values to each term in our  $PC_1$  equation (see Equation A1). Thus, we calculate a constant,  $C_1$ , from the weighted fractions of  $\bar{x}_i/s_{x_i}$  for each variable,  $x_i$ . Explicitly,

$$C_1 = \sum_i \frac{a_{i,1}(\bar{x}_i)}{s_{x_i}} \quad (13)$$

and thus,  $PC_1$  can be expressed in its uncentered form:

$$\begin{aligned} PC_{1,uncentered} &= PC_1 + C_1 \\ &= PC_1 + 6.616 \end{aligned} \quad (14)$$

It is also worth noting that, since the PCs define directions in space, the scaling is arbitrary. We can use the tight correlation with  $W(\lambda 5797)$  to rescale  $PC_1$  such that as the strength of  $\lambda 5797$  doubles, so too does  $N_{DIB}$ . The result is Equation 15.

$$\begin{aligned} N_{DIB} &= 0.136(PC_{1,uncentered}) \\ &= 0.136(PC_1) + 0.900 \end{aligned} \quad (15)$$

Since  $PC_1$  is not directly measurable, we provide an alternate expression in terms of  $W(\lambda 5797)$ .

$$N_{DIB} \approx 0.0185W(\lambda 5797) + 0.19 \quad (16)$$

The constant in Equation 16 indicates that some DIB carriers are present before the  $\lambda 5797$  carriers start to form. This result is further discussed in Section 5.

#### 4.2. $PC_2$ : strength of the ambient radiation field

$PC_2$  represents the second largest source of variations in the DIB strengths and line of sight parameters in our sample, and must be a parameter that is uncorrelated with  $PC_1$ . The largest contribution to  $PC_2$  comes from

$W(\lambda 5797)/W(\lambda 5780)$  (see Fig. 7 and Eq. (A2)) which correlates strongly (but negatively) with  $PC_2$ . The second largest projection onto  $PC_2$  is  $W(\lambda 6284)$ , in the positive direction. Note that some of our original parameters are hardly influenced by  $PC_2$ , most notably  $W(\lambda 5797)$ ,  $W(\lambda 6379)$ ,  $N(H)$  and  $E(B-V)$ .

These correlations (and lack thereof) are valuable clues to infer the physical quantity that is represented by  $PC_2$ . We start by noting that  $W(\lambda 5797)/W(\lambda 5780)$  is a variable that is clearly related to physical conditions, and has in particular been linked to the amount of exposure to the ambient UV radiation field (see e.g. Krelowski & Sneden 1995; Herbig 1995; Cami et al. 1997; Sonnentrucker et al. 1997; Krelowski et al. 1999; Vuong & Foing 2000; Cox et al. 2006; Friedman et al. 2011; Vos et al. 2011). More specifically, these studies have shown that the  $\lambda 5780$  DIB is weak in shielded environments, but becomes significantly stronger with increasing UV exposure in diffuse clouds; the  $\lambda 5797$  DIB on the other hand is already strong in shielded environments, and is largely indifferent to UV exposure in most diffuse clouds. In the harshest environments (e.g. Orion), both DIBs are very weak or absent. This different dependence on the radiation field then results in the more sheltered  $\zeta$ -type clouds having a high  $W(\lambda 5797)/W(\lambda 5780)$  ratio and the more exposed  $\sigma$ -type cloud environments having a low ratio. This dependence on the radiation field is also responsible for the so-called “skin effect” – the fact that DIB carriers seem to be concentrated near the surfaces of clouds (see e.g. Adamson et al. 1991; Herbig 1995; Cami et al. 1997; Vos et al. 2011).

One way such behaviour can be understood is by considering that the DIB carriers are molecules in a specific ionization state, and the changing DIB strengths reflect the molecules undergoing ionization with the carrier of the  $\lambda 5797$  having a lower ionization potential than the  $\lambda 5780$  DIB carrier (Cami et al. 1997; Sonnentrucker et al. 1997). Note that the charge balance is really determined by the interplay between ionization and recombination, and thus not only depends on the radiation field, but also on the density. Similarly, the DIB behaviour can be explained by the DIB carriers being a specific hydrogenation or protonation state (Vuong & Foing 2000; Le Page et al. 2001, 2003); also in this case, the density plays a role.

Thus, the  $W(\lambda 5797)/W(\lambda 5780)$  ratio certainly traces cloud depth, which encompasses UV exposure, density, and possibly also temperature. Given the strong correlation with  $PC_2$ ,  $PC_2$  could thus trace directly the

strength of the ambient radiation field  $G_0^3$ , or alternatively it could be a measure for  $G_0/n_e$  and reflect the balance between ionization and recombination (or similarly between hydrogenation and de-hydrogenation).

We dismiss this second interpretation for two reasons. First, Gnaciński et al. (2007) showed that several of the DIBs included in our sample are independent of  $n_e$ , contrary to the large variations we see among their  $PC_2$  values. Furthermore, Kos & Zwitter (2013) found no differences in  $n_e$  values between  $\sigma$  and  $\zeta$  sightlines, despite  $W(\lambda 5797)/W(\lambda 5780)$  having the largest projection on  $PC_2$ . Thus, we conclude that  $PC_2$  must be tracing changes in  $G_0$  only, although of course this is in turn can be determined by the cloud depth.

If we dichotomize sightlines into  $\sigma$  and  $\zeta$  clouds, this becomes more clear. Referring again to Figure 7, the negative projections onto  $PC_2$  are all indicators of a sheltered  $\zeta$  environment: We see a strong  $W(\lambda 5797)/W(\lambda 5780)$ , hydrogen in molecular form (i.e. large  $f(H_2)$ ), and the existence of  $C_2$  (demonstrated by the four  $C_2$ -DIBs –  $\lambda 5769$ ,  $\lambda 5546$ ,  $\lambda 5513$ , and  $\lambda 4964$ ; Thorburn et al. 2003). In contrast, positive  $PC_2$ -projections are consistent with more exposed conditions: we observe hydrogen in neutral form, as well as a strong projection from  $\lambda \lambda 5780$ ,  $4428$  and  $6284$  – DIBs known to correlate well with  $H\ I$  (Herbig 1993). Finally, both  $N(H)$  and  $E(B-V)$  contribute almost nothing to  $PC_2$ . Again, this is consistent with our interpretation since both variables contribute to the amount of material but are generally unaffected by changes in  $G_0$ . Note that  $\lambda 5797$  and  $\lambda 6379$  exhibit the same behavior, as was found previously (Cami et al. 1997; Kos & Zwitter 2013; Lan et al. 2015). The case of  $\lambda 5797$  is especially interesting. Although we use  $W(\lambda 5797)/W(\lambda 5780)$  to differentiate between regions of high and low UV exposure, these changes almost exclusively reflect the  $\lambda 5780$  carrier. Thus, while the carrier of  $\lambda 5780$  is sensitive to changes in  $G_0$ ,  $\lambda 5797$  survives at a steady strength per unit reddening over a wide range of conditions. This indifference to physical conditions is essentially why this DIB is a good tracer for the amount of DIB carrier material.

Comparing the three separate panels of Figure 7, we once again see that the PCA results from the line of sight parameters (bottom left) are quite different.  $N(H)$ , in particular, has a strong, positive projection on  $PC_2$ ,

<sup>3</sup>  $G_0$  is a convenient measure for the strength of the FUV field that is often used in the context of Photo-Dissociation Regions (PDRs; Tielens & Hollenbach 1985). It is the FUV field measured in units of the equivalent Habing (1968) flux of  $1.6 \times 10^{-3}$  ergs  $cm^{-2} s^{-1}$  appropriate to the average interstellar medium.

making it clear that  $PC_2$  traces something other than  $G_0$  in this case. This is not surprising since  $PC_2$  necessarily depends on  $PC_1$ , and  $PC_1$  was different between the full PCA and the PCA for the line of sight parameters only. The PCA results using only the DIBs, however, are nearly identical to those obtained with all 23 variables. This suggests that changes in  $PC_2$  – that is to say, changes in  $G_0$  – are directly observable through the relative strengths of DIBs.

For the sake of clarity, we further formalize our association of  $PC_2$  with the radiation field by defining a new parameter  $G_{DIB}$ . Ideally, we would like to scale this parameter to represent actual  $G_0$  values; however, without knowledge of specific  $G_0$  values for our targets, we cannot do this. Therefore, we leave  $PC_2$  as it is, and simply assign the name  $G_{DIB}$ :

$$G_{DIB} = PC_2 \quad (17)$$

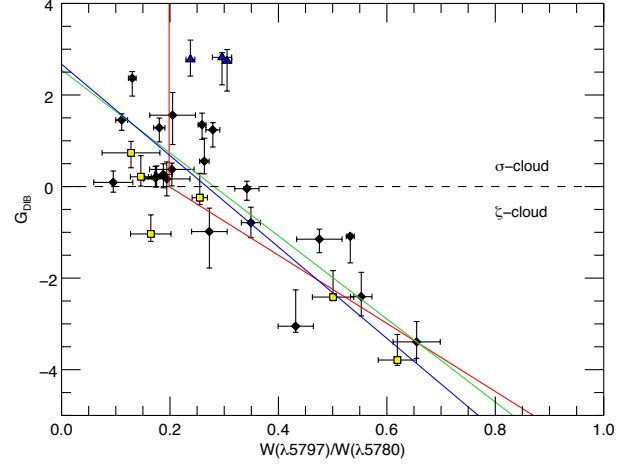
Since  $W(\lambda 5797)/W(\lambda 5780)$  correlates most strongly with  $PC_2$ , we will approximate  $G_{DIB}$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . Figure 9 shows  $G_{DIB}$  as a function of the  $W(\lambda 5797)$  to  $W(\lambda 5780)$  strength ratio. Fitting a straight line to the data gives the following relation with  $r = -0.716$  (shown in green in Fig. 9):

$$G_{DIB} \approx -9.07[W(\lambda 5797)/W(\lambda 5780)] + 2.55 \quad (18)$$

Arguably, a linear fit is not appropriate since there are three sightlines that strongly deviate from this relation – HD 15137, HD 198478, and HD 209975. However, the interstellar features for these sightlines are quite broad, suggesting that there may be more than one cloud in the line of sight that we are unable to resolve. If this is the case, then we cannot expect them to follow the same trend as true single clouds. Ignoring these three sightlines in our linear fit, we obtain (with  $r = -0.832$ ; shown in blue in Fig. 9):

$$G_{DIB} = -9.98[W(\lambda 5797)/W(\lambda 5780)] + 2.67 \quad (19)$$

Alternatively, these sightlines are in fact single clouds and represent a true trend in the data. With this interpretation, the linear dependence breaks down for positive  $G_{DIB}$  values (i.e.,  $\sigma$ -type clouds) and  $G_{DIB}$  increases independently of  $W(\lambda 5797)/W(\lambda 5780)$ . As previously discussed,  $W(\lambda 5797)$  is independent of  $PC_2$ , while  $W(\lambda 5780)$  shows a strong dependence; hence, this trend is most easily explained as an ionization effect of the  $\lambda 5780$  carrier. We propose that carrier of  $\lambda 5780$  is a cation or a dehydrogenated species (see e.g. Cami et al. 1997; Sonnentrucker et al. 1997). As  $G_0$  increases,



**Figure 9.**  $G_{DIB}$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . Be stars are indicated by yellow squares; three possible outliers (HD 15137, HD 198478, and HD 209975; located at  $G_{DIB} \approx 3$ ) are shown as blue triangles. We show the three fits discussed in the text: Eq. (18) in green; the fit without outliers from Eq. (19) in blue and the piece-wise function from Eq. (20) in red. See text for details.

more of these molecules become ionized/dehydrogenated so  $W(\lambda 5780)$  becomes stronger. Finally, when a certain  $G_0$  is reached, we enter the  $\sigma$ -cloud regime, where the cation/dehydrogenated state dominates. Hence,  $W(\lambda 5780)$  is no longer dependent on the environment, and  $\lambda 5797$  and  $\lambda 5780$  exist in approximately equal proportions. We can model this using a piece-wise function. For  $\sigma$  clouds,  $W(\lambda 5797)/W(\lambda 5780)$  has a mean value of  $0.198 \pm 0.064$ , with a maximum  $W(\lambda 5797)/W(\lambda 5780)$  value around 0.3. For  $\zeta$  clouds, the linear relation holds. We fit a line to the negative  $G_{DIB}$  values, forcing the intercept through the  $\sigma$  cloud mean using the MPFITEXY<sup>4</sup> routine (Williams et al. 2010) based on the MPFIT package (Markwardt 2009). We obtain a correlation coefficient of  $-0.75$  for  $\zeta$  clouds. The resulting equation is:

$$G_{DIB} \approx \begin{cases} \text{undefined,} & \text{if } \frac{W(\lambda 5797)}{W(\lambda 5780)} \lesssim 0.3 \\ -8.44\left(\frac{W(\lambda 5797)}{W(\lambda 5780)}\right) + 1.94, & \text{if } \frac{W(\lambda 5797)}{W(\lambda 5780)} \gtrsim 0.3 \end{cases} \quad (20)$$

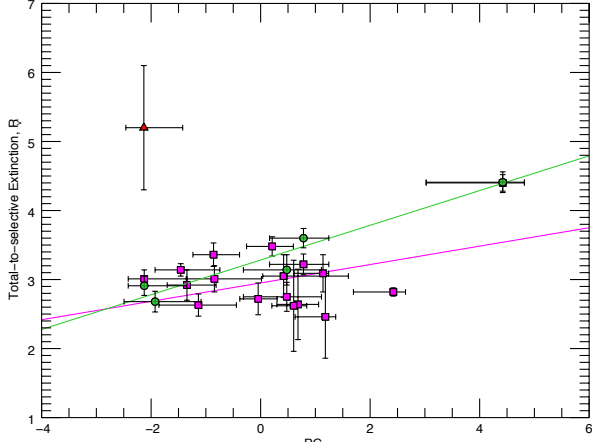
From this, it is clear that  $W(\lambda 5797)/W(\lambda 5780)$  is a suitable measure for  $G_0$  within  $\zeta$  clouds, whereas for  $\sigma$  clouds, it may or may not be appropriate.

#### 4.3. $PC_3$

<sup>4</sup> This is a functional linear regression taking into account errors in both  $x$  and  $y$  values.







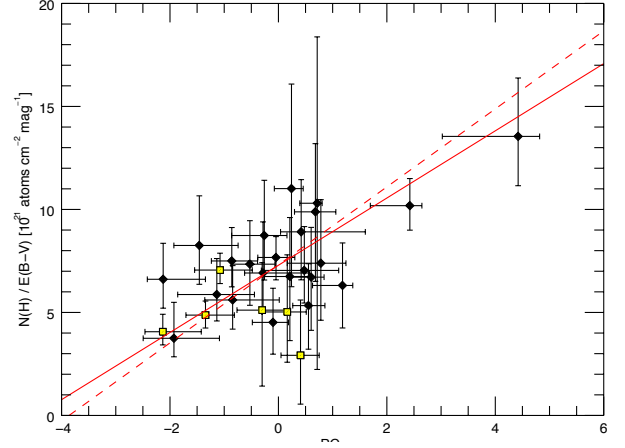
**Figure 11.** Total-to-selective extinction ratio,  $R_v$ , as a function of  $PC_3$ .  $R_v$  values are taken from: Wegner (2003, magenta squares); Fitzpatrick & Massa (2007, green circles) and Andersson & Potter (2007, red triangles). Note that there is some overlap between the Wegner (2003) and Fitzpatrick & Massa (2007) targets. The magenta and green lines are straight line fits (using IDL’s FITEXY routine) to the corresponding data sets by Wegner (2003) and Fitzpatrick & Massa (2007) respectively. We did not merge both data sets because of some discrepancies between the reported  $R_v$  values in both sets. Note that the value found for HD 110432 by Andersson & Potter (2007) deviates significantly from the mean trend.

indicate larger dust-to-gas ratios. Both  $f(H_2)$  and  $N(H_2)$  have large, negative projections, which again is consistent with a larger dust grain surface area and therefore a higher dust-to-gas ratio.  $E(B-V)$  also increases in this direction, with similar reasoning. Finally, we see a large projection from  $F_*$ . Noting that the large grains toward HD 147933 come from grain coagulation rather than elemental depletions (Jura 1980), this too is consistent: If there is more dust, then there is more material onto which gas may deplete. Both  $N(H)$  and  $N(H\ I)$  have positive projections, consistent with higher ratios of gas.  $N(H\ I)$ , in particular, has a large projection since a large quantity of dust is required to transform  $H\ I$  into  $H_2$ .

Figure 12 shows the ratio of hydrogen to color excess,  $N(H)/E(B-V)$ , plotted against  $PC_3$ . The general trend is apparent, and can be represented by (using IDL’s FITEXY routine):

$$\frac{N(H) \times 10^{-21}}{E(B-V)} = (7.29 \pm 0.23) + PC_3 \times (1.63 \pm 0.26) \quad (21)$$

However, the trendline appears to be influenced by the two outliers – HD 147933 and HD 207198 – and we therefore performed another fit excluding these two data



**Figure 12.** The total column of hydrogen per unit reddening as a function of  $PC_3$ . Yellow squares are used to indicate Be stars, which mostly fall below the trend line. The solid red line represents a straight line fit through all data points (Eq. 21; the dashed red line is the same excluding the two outliers (Eq. 22). Note that HD 143275 is not included in this plot since it has an  $E(B-V)$  value of zero.

points; this yields similar results:

$$\frac{N(H) \times 10^{-21}}{E(B-V)} = (7.31 \pm 0.27) + PC_3 \times (1.89 \pm 0.26) \quad (22)$$

There is quite a bit of scatter around this trend though. This is somewhat expected since  $N(H)/E(B-V)$  does not measure the gas-to-dust ratio directly, as  $E(B-V)$  measures the degree of reddening but not the dust mass. Moreover, our values for  $N(H)$  do not include any hydrogen that may exist in ionized form, while both HD 36822 and HD 36861 are known to lie near  $H\ II$  regions. Finally,  $E(B-V)$  values tend to be over-estimated for Be stars (see Section 2.1), resulting in smaller-than-expected  $N(H)/E(B-V)$  values. Be stars are differentiated in Figure 12 by square data points.

Although we cannot conclusively state what precisely  $PC_3$  is tracing, it seems clear that it is tracing a property of the dust and its relation with the gas. We thus favour the gas-to-dust ratio interpretation here.

Once again, we define a new parameter, GTD, describing the observed changes along  $PC_3$ . Bohlin et al. (1978) showed that the mean interstellar value of  $N(H)/E(B-V)$  is  $5.8 \times 10^{21}$  atoms  $cm^{-2}$   $mag^{-1}$ . We rescale  $PC_3$  such that GTD gives the ratio of gas to dust with respect to the mean interstellar value: A GTD value of 1 corresponds to the mean interstellar value of  $N(H)/E(B-V)$ , and HD 147933 has a GTD value of 2.7. GTD is therefore expressed as:

$$GTD = 0.318(PC_3) + 1.29 \quad (23)$$

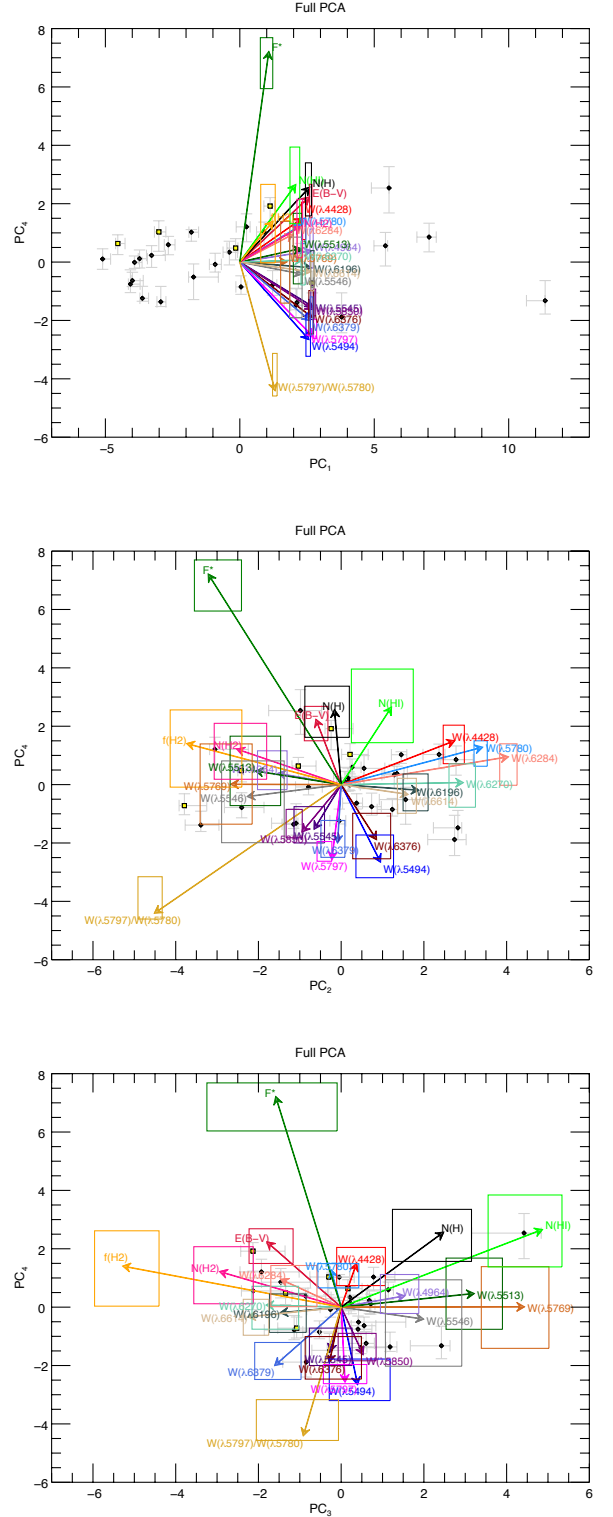
4.4.  $PC_4$ 

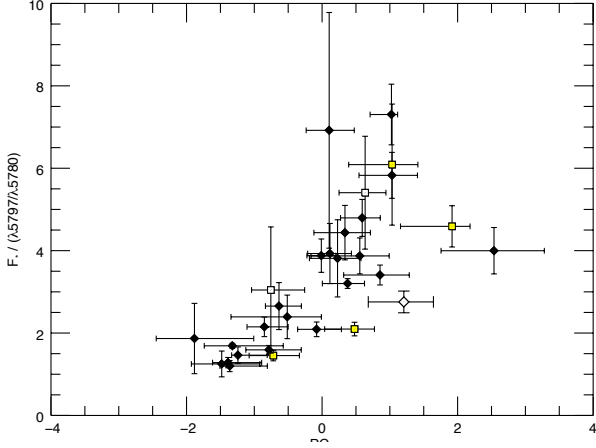
Figure 13 shows the  $PC_4$  biplots with  $PC_1$  (top),  $PC_2$  (middle), and  $PC_3$  (bottom). It is difficult to physically interpret  $PC_4$  due to the large uncertainties on the projections; however, it is clear that the strongest projections come from  $F_\star$  and  $W(\lambda 5797)/W(\lambda 5780)$ , and thus,  $PC_4$  traces some difference between these two variables. In the previous PCs,  $F_\star$  and  $W(\lambda 5797)/W(\lambda 5780)$  increased in the same direction, suggesting that depletions generally increase with cloud depth. In terms of  $PC_4$ , however, these two variables increase in opposite directions. Consequently, a possible interpretation is that  $PC_4$  traces the deviations from the mean relation between these two variables, although the physical cause for these deviations is unknown. Figure 14 shows the ratio of these variables plotted as a function of  $PC_4$ .

## 5. DISCUSSION

The key result obtained through this exercise is that a variety of DIBs in single cloud lines of sight representing very different environments can be described remarkably well by just four parameters – a surprising fact considering the large number of DIBs and the lack of strong correlations among them. This implies that differences in interstellar environments are not driven by the DIBs themselves, but rather, the DIBs react to a limited number of parameters that dictate ISM conditions. Each DIB has a unique response to these conditions, and thus we see a unique DIB spectrum in all lines of sight. Note that also Lan et al. (2015) could reduce much of the variation in the average DIB properties to just two parameters.

DIB strength is predominantly determined by the parameter  $N_{\text{DIB}}$ . In Figure 15, we show the equivalent width of each DIB included in our sample as a function of  $N_{\text{DIB}}$ . From this figure, we can gain two important insights. The first is that we can see how capable  $N_{\text{DIB}}$  is of predicting the strength of each DIB. Those DIBs that correlate very strongly (e.g.,  $\lambda 5797$ ,  $\lambda 5850$ ,  $\lambda 5494$ ) can be predicted by a single parameter – that is to say,  $N_{\text{DIB}}$  is singularly capable of describing their observed strengths. For DIBs that correlate less strongly (e.g.,  $\lambda 5769$ ,  $\lambda 6284$ ,  $\lambda 5513$ ), there are other parameters (such as  $G_{\text{DIB}}$  and  $G_{\text{T D}}$ ) playing a significant role. The second insight gained from Figure 15 is that we can infer the order in which DIBs form. For example, if the line of best fit intersects the x-axis at a small value (e.g.,  $\lambda 6284$ ,  $\lambda 4428$ ,  $\lambda 5780$ ,  $\lambda 6270$ ) then these carriers start to form when  $N_{\text{DIB}}$ , the collective level of DIB-producing material, is still small. In other words, these carriers form before the others. On the other hand, DIBs like  $\lambda 5850$ ,  $\lambda 4964$ , and  $\lambda 5513$  start forming once the aforementioned





**Figure 14.**  $PC_4$  traces some difference between  $F_*$  and  $W(\lambda 5797)/W(\lambda 5780)$ . Here,  $r = -0.769$ . Squares (white and yellow) indicate Be stars. White data points (squares and diamonds) indicate synthetically-derived  $F_*$  values.

DIB carriers are already present in appreciable amounts. These differences might reflect the molecular complexity of the carriers (i.e., simple species form at lower  $N_{\text{DIB}}$  values and may act as precursors for those species forming at higher  $N_{\text{DIB}}$  values) or ionization potentials. It should be noted that negative  $N_{\text{DIB}}$  intercepts are permissible, since  $N_{\text{DIB}}$  acts like the column density of the “average” DIB carrier.

The second most important parameter for determining DIB strength is the amount of UV radiation. To specifically investigate how DIBs respond to changes in UV exposure, we must first normalize DIB strengths to the amount of material. We illustrate this in three different ways in Figure 16 for three DIBs:  $\lambda 6376$  (whose variation is primarily determined by  $N_{\text{DIB}}$ ),  $\lambda 4964$  (a  $C_2$ -DIB), and  $\lambda 6270$  (which has quite a bit of variability unaccounted for by  $N_{\text{DIB}}$ ). Similar plots for our full set of DIBs are shown in Fig. 17 in the appendix. In the leftmost plots, we employ the common practice of normalizing DIBs to  $E(B-V)$ , and plot  $W(\lambda 5797)/W(\lambda 5780)$  (the variable that best correlates with  $G_{\text{DIB}}$ ) on the x-axis. In the center plots, we instead normalize DIBs to  $W(\lambda 5797)$  – the variable which most strongly correlates with  $N_{\text{DIB}}$ . Finally, in the rightmost plots, DIBs are normalized to  $N_{\text{DIB}}$  directly and are plotted as a function of  $G_{\text{DIB}}$ .

A comparison of the left and center plots reveals the consequences of normalizing DIBs to  $E(B-V)$  versus  $W(\lambda 5797)$ . The latter situation tends to yield clearer trends compared to the former. This is most apparent for  $\lambda 6270$ , which exhibits clear environmental behavior. However, even when the main trend is less discernible

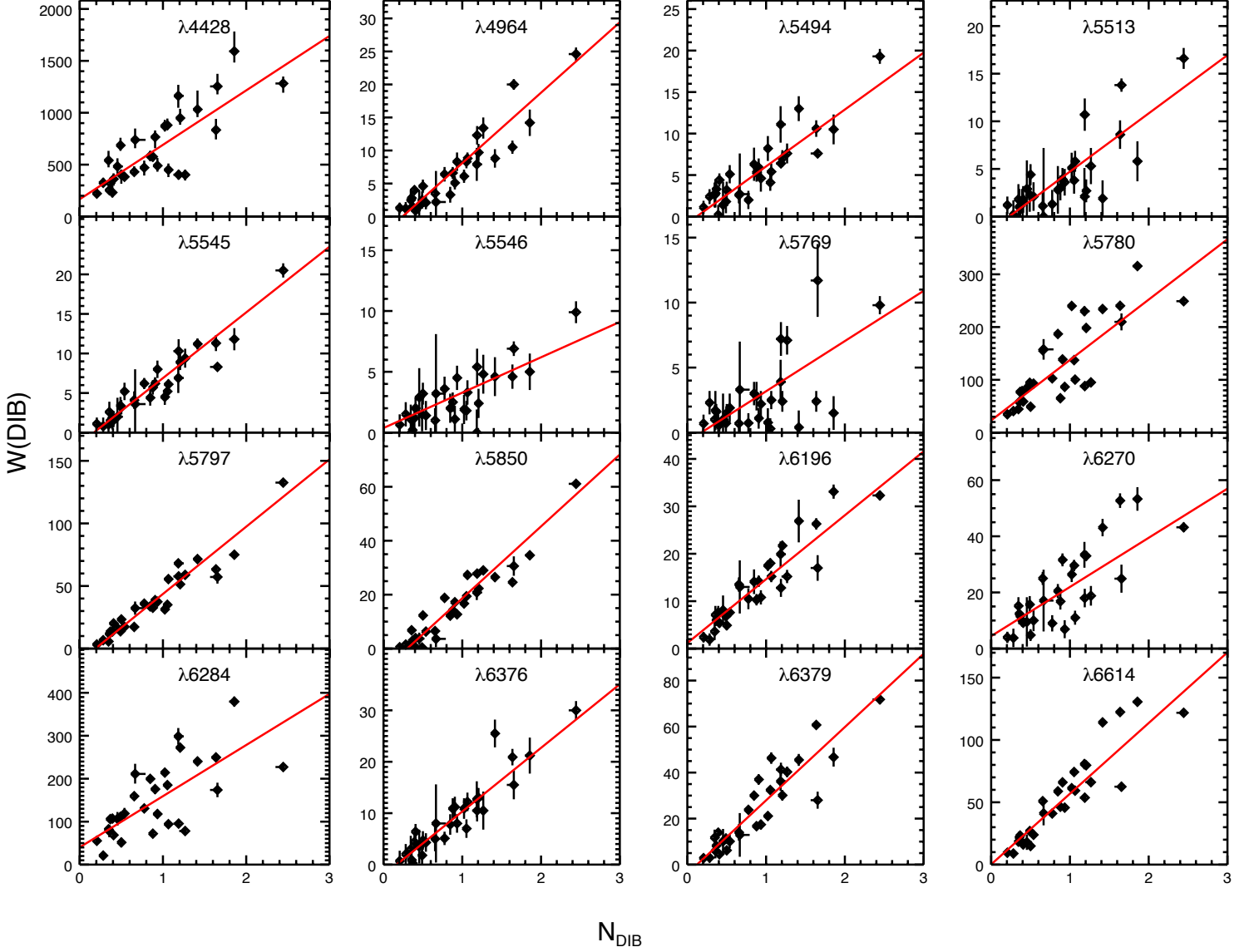
(e.g., for  $\lambda 6376$ ), a reduced amount of scatter is still noted.

The plots on the right provide a clearer picture of DIB environmental behavior, although they have the disadvantage of not being reproducible without individually measuring the variables corresponding to the 23 terms in Equation A2 or else performing PCA. Note that the x-axis is reversed in these plots, compared to the left and center ones (i.e.,  $PC_1$  increases for decreasing  $W(\lambda 5797)/W(\lambda 5780)$  values).  $\lambda 6270$  has a positive slope, suggesting that it is an ionized or dehydrogenated species. As  $G_{\text{DIB}}$  increases, more carriers become ionized dehydrogenated and this DIB becomes stronger. It makes sense that  $\lambda 6270$  would be strongly influenced by  $G_{\text{DIB}}$  since, referring to Figure 15,  $N_{\text{DIB}}$  is unable to account for a sizeable portion of this DIB’s variation. The remaining variation must be driven by the subsequent PCs and hence,  $G_{\text{DIB}}$  shows a strong trend for  $\lambda 6270$ .  $\lambda 4964$ , on the other hand, displays the opposite behavior: As  $G_{\text{DIB}}$  increases, the strength of  $\lambda 4964$  decreases, suggesting that the carrier prefers more sheltered environments (e.g., lower  $G_0$ , higher densities, and cooler temperatures). For  $\lambda 6376$ , the trend with  $G_{\text{DIB}}$  is not particularly strong, suggesting that this feature does not systematically vary with  $G_0$  over the range of UV exposures covered in this study. Incidentally, this DIB has a very small projection on  $PC_2$  (see Figure 7).

The physical interpretation for  $PC_3$  is less certain. Because of this fact, we are limited in our ability to infer characteristics of the DIB carriers based on their behavior with respect to  $PC_3$ . It seems clear though that  $PC_3$  is tracing a property of the dust or its relation with the gas (such as e.g. the gas-to-dust ratio), suggesting that dust may play a role in the formation and/or excitation of some of the DIB carriers.

## 6. CONCLUSIONS

We performed a principal component analysis on a selection of DIBs and line of sight parameters measured for single-cloud sightlines. We found that the majority of DIB variations can be attributed to four parameters. The variable that most strongly determines DIB strength is the amount of DIB-producing material in the line of sight, a parameter that is traced extremely well by  $W(\lambda 5797)$ . The second most important parameter is the level of UV exposure, which is reasonably well-approximated by  $W(\lambda 5797)/W(\lambda 5780)$  in  $\zeta$  sightlines.  $W(\lambda 5797)$  is unaffected by this parameter, and therefore changes in  $W(\lambda 5797)/W(\lambda 5780)$  primarily reflect changes in the  $\lambda 5780$  carrier. The third is presumably related to the dust properties in the line of sight, with the gas-to-dust ratio being one possibility. Finally, the

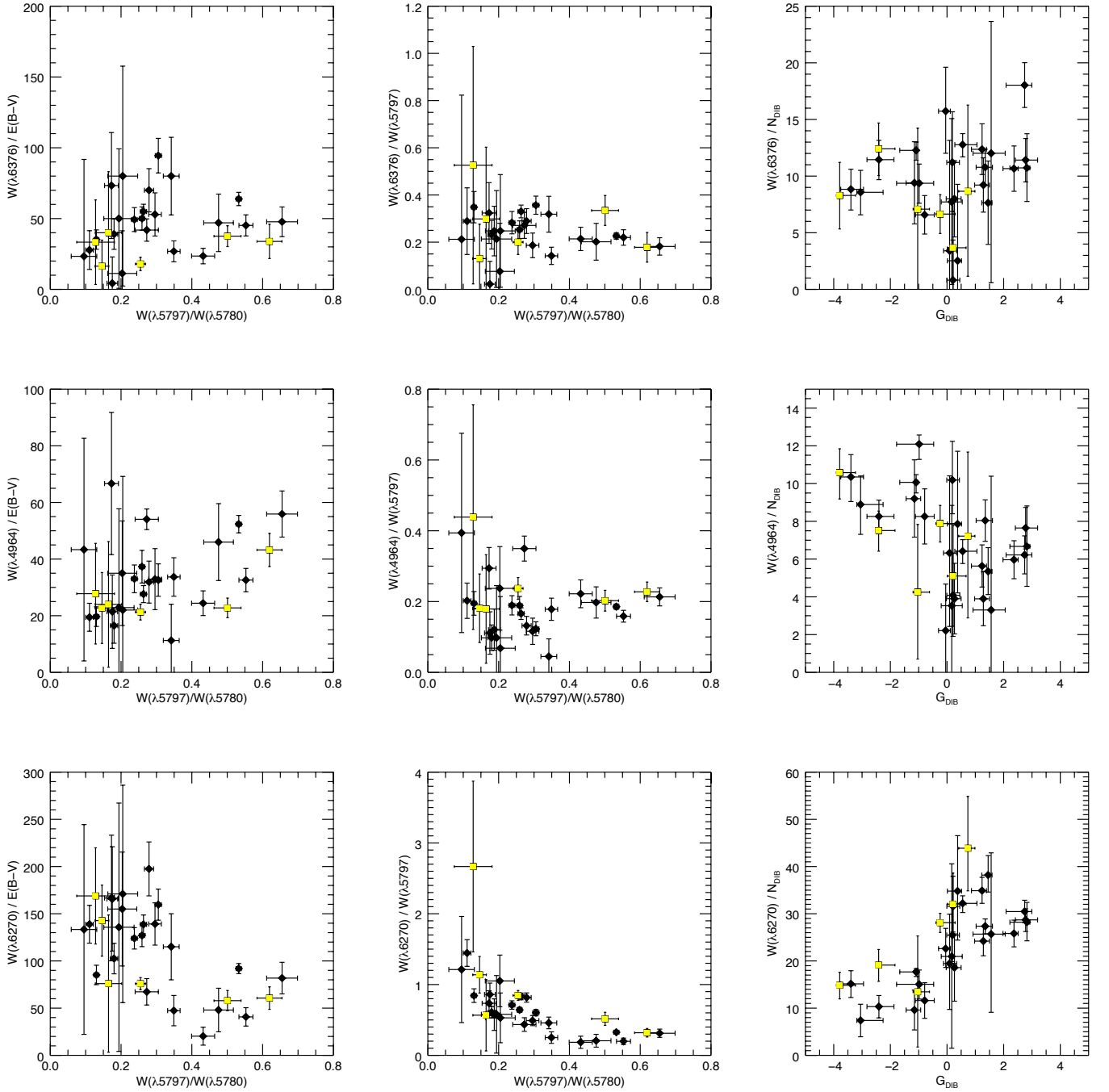


**Figure 15.** The strength of each DIB plotted against  $N_{\text{DIB}}$ . In each case, the line of best fit is obtained through linear least squares fitting using IDL's FITEXY routine. The correlation coefficients, from largest to smallest are  $W(\lambda 5797)$ : 0.957,  $W(\lambda 5850)$ : 0.952,  $W(\lambda 5494)$ : 0.948,  $W(\lambda 6376)$ : 0.936,  $W(\lambda 4964)$ : 0.931,  $W(\lambda 6196)$ : 0.928,  $W(\lambda 6614)$ : 0.920,  $W(\lambda 6379)$ : 0.912,  $W(\lambda 5494)$ : 0.897,  $W(\lambda 6270)$ : 0.811,  $W(\lambda 4428)$ : 0.811,  $W(\lambda 5780)$ : 0.807,  $W(\lambda 5546)$ : 0.806,  $W(\lambda 5513)$ : 0.795,  $W(\lambda 6284)$ : 0.691, and  $W(\lambda 5769)$ : 0.614. The x-intercepts from smallest to largest are:  $W(\lambda 6284)$ :  $-0.338 \pm 0.034$ ,  $W(\lambda 4428)$ :  $-0.317 \pm 0.038$ ,  $W(\lambda 6270)$ :  $-0.253 \pm 0.057$ ,  $W(\lambda 5780)$ :  $-0.202 \pm 0.018$ ,  $W(\lambda 5546)$ :  $-0.128 \pm 0.146$ ,  $W(\lambda 6196)$ :  $-0.087 \pm 0.036$ ,  $W(\lambda 6614)$ :  $-0.003 \pm 0.019$ ,  $W(\lambda 5494)$ :  $0.119 \pm 0.050$ ,  $W(\lambda 6379)$ :  $0.123 \pm 0.024$ ,  $W(\lambda 6376)$ :  $0.166 \pm 0.059$ ,  $W(\lambda 5545)$ :  $0.171 \pm 0.045$ ,  $W(\lambda 5769)$ :  $0.177 \pm 0.082$ ,  $W(\lambda 5797)$ :  $0.188 \pm 0.014$ ,  $W(\lambda 5513)$ :  $0.229 \pm 0.077$ ,  $W(\lambda 4964)$ :  $0.241 \pm 0.040$ , and  $W(\lambda 5850)$ :  $0.301 \pm 0.018$ .

fourth parameter is related to the depletions in the line of sight, although we do not offer a physical interpretation beyond this.

The work presented in this paper is only a first attempt at discriminating the different parameters that drive the variations in the DIBs and line of sight parameters. Our results show that there is great potential in this method to recognize the processes at play, and

to turn the DIBs themselves into powerful probes of the conditions in their surroundings. With the current measurement errors however, only the first three principal components can be somewhat confidently identified; the uncertainties on the remaining components are too large for a meaningful interpretation. At this point, it is not clear whether the uncertainties on these components is the result of intrinsic differences in the DIBs (e.g. dif-



**Figure 16.** The DIBs (top to bottom)  $\lambda 6376$ ,  $\lambda 4964$ , and  $\lambda 6270$  (*Left:*) normalized to  $E(B-V)$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ ; (*Centre:*) normalized to  $W(\lambda 5797)$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ ; (*Right:*) normalized to  $N_{\text{DIB}}$  as a function of  $G_{\text{DIB}}$ . Be stars are indicated by yellow squares. HD 143275 has been excluded from the left plots because it has an  $E(B-V)$  value of zero.



ferent carrier abundances), or whether it is the direct consequence of measurement errors. In future work, this will become clear when using a much larger sample of DIBs from high signal-to-noise observations of many more sightlines in the context of the European Southern Observatory Diffuse Interstellar Bands Large Exploration Survey (EDIBLES). With a larger number of observations, one could also consider a different PCA that can then result in “principal eigenspectra”. While such an exercise would be very interesting, it can only be expected to work if the different spectral sources (stellar, interstellar and telluric) can be reliably separated.

We thank the anonymous referee for valuable comments that improved the clarity of this manuscript. TE, JC, NHB, and AS acknowledge the support from the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery grant. This research has made use of NASA’s Astrophysics Data System and of the SIMBAD database, operated at CDS, Strasbourg, France.

*Software:* IDL; molecfit (Smette et al. 2015; Kausch et al. 2015); MPFITEXY (Williams et al. 2010); MPFIT package (Markwardt 2009)

*Facility:* VLT:Kueyen (UVES), OHP:1.93m (ELODIE)

## APPENDIX

### A. MEASUREMENTS

Below is an overview of all the DIB equivalent width measurements that are used in this paper. Note that the uncertainties on the measurements were obtained by using a Monte Carlo approach to simulate positioning the continuum. Table 5 lists the measurements for the  $\lambda\lambda 4428, 4964, 5494, 5513, 5545, 5546, 5769$  and  $5780$  DIBs; Table 6 lists the same for the  $\lambda\lambda 5797, 5850, 6196, 6270, 6284, 6376, 6379$  and  $6614$  DIBs.

### B. THE MAIN PCS

As is discussed in the main body of the paper, only the first four principal components can be interpreted in a meaningful way. While the method and the general formalism is presented in Sect. 3, it is useful and insightful to have the full expressions (i.e. the coefficients  $a$  in Eq. (1)) for the principal components in terms of the original variables as well as in terms of the standardized variables. Below are the equations for the leading four PCs in terms of the standardized variables ( $z_1, z_2, \dots, z_{23}$ ) and then re-expressed in terms of the original, measurable variables. See Sect. 3 for details.

$$\begin{aligned}
 PC_1 = & 0.208(z_1) + 0.239(z_2) + 0.230(z_3) + 0.203(z_4) + 0.243(z_5) + 0.206(z_6) + 0.157(z_7) \\
 & + 0.207(z_8) + 0.245(z_9) + 0.244(z_{10}) + 0.238(z_{11}) + 0.208(z_{12}) + 0.177(z_{13}) + 0.240(z_{14}) \\
 & + 0.234(z_{15}) + 0.236(z_{16}) + 0.232(z_{17}) + 0.186(z_{18}) + 0.203(z_{19}) + 0.232(z_{20}) + 0.110(z_{21}) \\
 & + 0.096(z_{22}) + 0.118(z_{23}) \\
 = & (5.92 \times 10^{-4})W(\lambda 4428) + (4.18 \times 10^{-2})W(\lambda 4964) + (5.56 \times 10^{-2})W(\lambda 5494) + (5.24 \times 10^{-2})W(\lambda 5513) \\
 & + (5.63 \times 10^{-2})W(\lambda 5545) + (9.36 \times 10^{-2})W(\lambda 5546) + (5.62 \times 10^{-2})W(\lambda 5769) + (2.69 \times 10^{-3})W(\lambda 5780) \\
 & + (8.83 \times 10^{-3})W(\lambda 5797) + (1.80 \times 10^{-2})W(\lambda 5850) + (2.85 \times 10^{-2})W(\lambda 6196) + (1.49 \times 10^{-2})W(\lambda 6270) \\
 & + (2.10 \times 10^{-3})W(\lambda 6284) + (3.24 \times 10^{-2})W(\lambda 6376) + (1.30 \times 10^{-2})W(\lambda 6379) + (6.70 \times 10^{-3})W(\lambda 6614) \\
 & + (1.75)E(B - V) + (2.03 \times 10^{-22})N(H_1) + (7.80 \times 10^{-22})N(H_2) + (1.95 \times 10^{-22})N(H) \\
 & + (4.72 \times 10^{-1})f(H_2) + (4.49 \times 10^{-1})F_\star + (7.50 \times 10^{-1})W(\lambda 5797)/W(\lambda 5780) - 6.62
 \end{aligned} \tag{A1}$$

**Table 5.** Equivalent widths and uncertainties of the  $\lambda\lambda 4428, 4964, 5494, 5513, 5545, 5546, 5769$  and  $5780$  DIBs.

Target	$\lambda 4428$	$\lambda 4964$	$\lambda 5494$	$\lambda 5513$	$\lambda 5545$	$\lambda 5546$	$\lambda 5769$	$\lambda 5780$
HD 15137	$1163 \pm \frac{106}{115}$	$7.9 \pm 2.5$	$11.1 \pm 2.2$	$2.1 \pm 3.0$	$6.9 \pm 1.9$	$0.0 \pm 1.9$	$3.9 \pm 1.7$	$230.1 \pm 9.1$
HD 22951	$471 \pm \frac{68}{76}$	$6.4 \pm 1.1$	$2.0 \pm 1.1$	$1.3 \pm 1.5$	$6.2 \pm 0.7$	$3.6 \pm 1.0$	$0.7 \pm 0.8$	$102.8 \pm 3.6$
HD 23180	$403 \pm \frac{45}{47}$	$12.3 \pm 1.4$	$6.4 \pm 0.2$	$10.7 \pm 1.7$	$10.3 \pm 1.5$	$5.4 \pm 1.5$	$7.2 \pm 1.3$	$88.1 \pm 5.0$
HD 23630	$325 \pm \frac{48}{39}$	$1.2 \pm 1.0$	$2.4 \pm 0.9$	$0.2 \pm 1.5$	$0.8 \pm 1.1$	$1.5 \pm 1.0$	$2.3 \pm 0.9$	$40.7 \pm 4.8$
HD 24398	$450 \pm \frac{61}{70}$	$8.8 \pm 0.9$	$5.4 \pm 1.0$	$5.8 \pm 1.1$	$6.1 \pm 0.6$	$3.3 \pm 1.0$	$2.5 \pm 0.7$	$100.4 \pm 2.7$
HD 24534	$402 \pm \frac{49}{55}$	$13.4 \pm 1.6$	$7.6 \pm 1.2$	$5.3 \pm 1.9$	$9.4 \pm 1.2$	$4.8 \pm 1.6$	$7.1 \pm 1.1$	$95.1 \pm 5.0$
HD 24760	$322 \pm \frac{41}{30}$	$1.5 \pm 0.8$	$3.3 \pm 0.8$	$1.1 \pm 1.0$	$1.5 \pm 0.9$	$0.2 \pm 0.8$	$1.6 \pm 0.6$	$77.0 \pm 3.4$
HD 24912	$949 \pm \frac{89}{65}$	$9.7 \pm 1.3$	$7.0 \pm 1.0$	$2.7 \pm 1.2$	$8.9 \pm 1.0$	$2.4 \pm 1.2$	$2.4 \pm 0.8$	$198.3 \pm 3.1$
HD 27778	$490 \pm \frac{74}{58}$	$8.3 \pm 1.4$	$4.6 \pm 1.6$	$3.6 \pm 1.4$	$8.0 \pm 1.1$	$4.5 \pm 1.0$	$2.2 \pm 1.0$	$86.6 \pm 4.6$
HD 35149	$254 \pm \frac{43}{38}$	$2.8 \pm 1.3$	$2.8 \pm 1.7$	$1.0 \pm 1.9$	$2.6 \pm 1.3$	$0.0 \pm 1.4$	$1.7 \pm 1.5$	$58.0 \pm 5.5$
HD 35715	$221 \pm \frac{47}{23}$	$1.3 \pm 0.8$	$1.1 \pm 0.8$	$1.2 \pm 0.9$	$1.1 \pm 0.8$	$0.7 \pm 0.9$	$0.7 \pm 0.7$	$34.6 \pm 3.6$
HD 36822	$483 \pm \frac{78}{69}$	$1.6 \pm 2.4$	$1.4 \pm 2.8$	$2.9 \pm 3.0$	$2.0 \pm 2.4$	$2.9 \pm 2.4$	$1.0 \pm 2.0$	$84.5 \pm 9.6$
HD 36861	$402 \pm \frac{49}{86}$	$4.6 \pm 1.0$	$3.2 \pm 1.0$	$4.4 \pm 1.1$	$3.2 \pm 0.9$	$3.2 \pm 0.9$	$1.5 \pm 0.7$	$49.0 \pm 3.5$
HD 40111	$739 \pm \frac{109}{81}$	$2.2 \pm 4.7$	$2.7 \pm 4.9$	$0.0 \pm 7.2$	$3.6 \pm 4.4$	$3.2 \pm 4.9$	$3.3 \pm 3.7$	$157.7 \pm 19.5$
HD 110432	$880 \pm \frac{64}{45}$	$8.3 \pm 1.0$	$4.1 \pm 1.0$	$3.8 \pm 1.4$	$5.2 \pm 1.0$	$1.8 \pm 0.8$	$0.3 \pm 0.8$	$137.3 \pm 3.7$
HD 143275	$383 \pm \frac{21}{12}$	$2.1 \pm 1.0$	$5.1 \pm 0.1$	$2.1 \pm 1.5$	$5.2 \pm 1.1$	$1.4 \pm 1.2$	$1.9 \pm 1.1$	$92.7 \pm 4.2$
HD 144217	$430 \pm \frac{54}{38}$	$3.5 \pm 0.8$	$2.6 \pm 1.0$	$1.1 \pm 1.6$	$4.1 \pm 1.1$	$1.0 \pm 1.0$	$0.7 \pm 1.1$	$156.0 \pm 4.9$
HD 145502	$583 \pm \frac{50}{48}$	$3.3 \pm 1.2$	$6.3 \pm 2.0$	$2.8 \pm 2.5$	$4.4 \pm 1.0$	$2.0 \pm 1.2$	$3.0 \pm 0.9$	$186.9 \pm 5.2$
HD 147165	$872 \pm \frac{50}{53}$	$6.1 \pm 1.0$	$8.2 \pm 1.5$	$5.1 \pm 1.6$	$4.5 \pm 1.0$	$1.9 \pm 1.2$	$0.8 \pm 1.1$	$240.0 \pm 4.2$
HD 147933	$1254 \pm \frac{121}{77}$	$20.0 \pm 0.8$	$7.6 \pm 0.5$	$13.8 \pm 0.7$	$8.3 \pm 0.5$	$6.9 \pm 0.6$	$11.7 \pm 2.8$	$209.8 \pm 16.1$
HD 149757	$576 \pm \frac{52}{47}$	$6.6 \pm 0.9$	$5.3 \pm 1.1$	$3.0 \pm 1.3$	$5.7 \pm 0.9$	$2.5 \pm 0.8$	$2.8 \pm 1.1$	$65.1 \pm 3.8$
HD 164284	$686 \pm \frac{73}{53}$	$2.5 \pm 1.3$	$1.8 \pm 1.4$	$2.3 \pm 1.8$	$3.4 \pm 1.1$	$1.5 \pm 1.4$	$0.7 \pm 1.0$	$94.4 \pm 4.4$
HD 170740	$834 \pm \frac{107}{91}$	$10.5 \pm 1.0$	$10.6 \pm 1.0$	$8.6 \pm 1.5$	$11.3 \pm 1.0$	$4.6 \pm 1.0$	$2.4 \pm 0.8$	$240.3 \pm 4.0$
HD 198478	$1592 \pm \frac{191}{108}$	$14.2 \pm 2.0$	$10.5 \pm 1.8$	$5.8 \pm 2.1$	$11.8 \pm 1.4$	$5.0 \pm 1.5$	$1.5 \pm 1.3$	$315.6 \pm 5.8$
HD 202904	$541 \pm \frac{92}{67}$	$2.5 \pm 1.5$	$2.6 \pm 1.5$	$1.8 \pm 1.6$	$1.0 \pm 1.0$	$1.1 \pm 1.2$	$1.0 \pm 1.1$	$44.5 \pm 4.6$
HD 207198	$1282 \pm \frac{67}{89}$	$24.6 \pm 1.0$	$19.3 \pm 0.9$	$16.6 \pm 1.1$	$20.5 \pm 0.9$	$9.9 \pm 0.9$	$9.8 \pm 0.7$	$249.0 \pm 2.8$
HD 209975	$1032 \pm \frac{182}{74}$	$8.8 \pm 1.4$	$13.0 \pm 1.5$	$1.9 \pm 1.9$	$11.2 \pm 0.7$	$4.6 \pm 1.6$	$0.4 \pm 1.3$	$234.2 \pm 4.7$
HD 214680	$361 \pm \frac{55}{64}$	$0.9 \pm 1.0$	$4.4 \pm 0.8$	$1.8 \pm 1.4$	$1.7 \pm 1.0$	$2.0 \pm 0.9$	$0.6 \pm 0.5$	$58.8 \pm 2.8$
HD 214993	$232 \pm \frac{63}{47}$	$4.0 \pm 0.7$	$0.2 \pm 1.2$	$1.6 \pm 1.3$	$1.4 \pm 1.0$	$1.2 \pm 0.9$	$0.6 \pm 0.8$	$78.6 \pm 4.8$
HD 218376	$766 \pm \frac{64}{108}$	$5.1 \pm 1.0$	$5.9 \pm 1.1$	$3.9 \pm 1.2$	$6.2 \pm 0.8$	$1.1 \pm 1.1$	$1.1 \pm 0.8$	$138.7 \pm 4.4$

NOTE—All measurements in mÅ.

**Table 6.** Equivalent widths and uncertainties of the  $\lambda\lambda 5797, 5850, 6196, 6270, 6284, 6376, 6379$  and  $6614$  DIBs.

Target	$\lambda 5797$	$\lambda 5850$	$\lambda 6196$	$\lambda 6270$	$\lambda 6284$	$\lambda 6376$	$\lambda 6379$	$\lambda 6614$
HD 15137	$68.1 \pm 3.1$	$20.8 \pm 2.8$	$19.9 \pm 2.7$	$33.4 \pm 4.6$	$298.6 \pm 19.4$	$12.7 \pm 3.4$	$36.2 \pm 4.2$	$80.6 \pm 4.1$
HD 22951	$35.9 \pm 1.3$	$18.8 \pm 1.2$	$10.5 \pm 2.2$	$9.0 \pm 2.9$	$130.8 \pm 8.5$	$5.1 \pm 1.3$	$23.8 \pm 1.5$	$41.0 \pm 2.1$
HD 23180	$57.7 \pm 2.0$	$27.8 \pm 1.3$	$12.8 \pm 1.9$	$18.0 \pm 3.3$	$95.4 \pm 9.4$	$10.5 \pm 2.1$	$41.3 \pm 3.0$	$53.7 \pm 3.4$
HD 23630	$6.7 \pm 1.3$	$1.5 \pm 1.0$	$1.9 \pm 1.3$	$3.8 \pm 3.3$	$21.0 \pm 7.7$	$2.0 \pm 2.0$	$3.0 \pm 2.1$	$8.9 \pm 2.8$
HD 24398	$55.5 \pm 1.3$	$27.3 \pm 1.1$	$15.2 \pm 1.2$	$11.0 \pm 2.5$	$94.1 \pm 6.7$	$12.2 \pm 1.8$	$46.3 \pm 2.5$	$59.3 \pm 1.9$
HD 24534	$58.9 \pm 1.3$	$29.0 \pm 1.4$	$15.2 \pm 1.4$	$18.8 \pm 3.5$	$78.2 \pm 8.2$	$10.5 \pm 3.7$	$40.3 \pm 2.3$	$66.1 \pm 2.4$
HD 24760	$13.5 \pm 1.0$	$2.9 \pm 0.8$	$6.0 \pm 1.2$	$11.6 \pm 2.0$	$105.9 \pm 5.5$	$0.3 \pm 1.3$	$8.2 \pm 1.5$	$23.3 \pm 2.1$
HD 24912	$51.4 \pm 1.2$	$22.3 \pm 1.7$	$21.7 \pm 1.0$	$33.0 \pm 1.7$	$272.4 \pm 9.6$	$13.0 \pm 1.9$	$30.1 \pm 2.3$	$79.7 \pm 1.8$
HD 27778	$37.4 \pm 2.0$	$12.7 \pm 1.3$	$10.8 \pm 1.5$	$6.9 \pm 3.2$	$117.8 \pm 10.2$	$8.0 \pm 1.8$	$17.4 \pm 2.1$	$45.7 \pm 2.7$
HD 35149	$11.8 \pm 2.1$	$6.8 \pm 1.3$	$7.1 \pm 1.9$	$12.4 \pm 3.7$	$78.0 \pm 14.4$	$0.9 \pm 2.4$	$6.0 \pm 3.3$	$21.9 \pm 4.6$
HD 35715	$3.3 \pm 1.2$	$0.5 \pm 0.7$	$2.4 \pm 1.1$	$4.0 \pm 2.0$	$55.4 \pm 8.4$	$0.7 \pm 2.0$	$2.8 \pm 1.9$	$9.5 \pm 1.9$
HD 36822	$16.4 \pm 3.1$	$3.7 \pm 2.2$	$8.1 \pm 3.1$	$9.5 \pm 8.8$	$106.6 \pm 15.9$	$3.5 \pm 3.3$	$10.1 \pm 5.4$	$18.0 \pm 6.2$
HD 36861	$23.3 \pm 1.2$	$12.3 \pm 0.8$	$4.9 \pm 1.0$	$4.8 \pm 2.1$	$51.6 \pm 10.8$	$4.7 \pm 1.8$	$6.2 \pm 1.4$	$14.9 \pm 1.8$
HD 40111	$32.3 \pm 5.3$	$3.6 \pm 3.1$	$13.0 \pm 5.6$	$17.1 \pm 1.0$	$211.1 \pm 22.5$	$8.0 \pm 7.6$	$12.9 \pm 9.5$	$41.1 \pm 9.6$
HD 110432	$35.0 \pm 1.7$	$19.4 \pm 1.0$	$18.0 \pm 1.0$	$29.6 \pm 2.0$	$185.1 \pm 5.1$	$7.0 \pm 1.8$	$32.4 \pm 1.8$	$74.3 \pm 2.1$
HD 143275	$17.4 \pm 1.3$	$6.3 \pm 1.1$	$7.6 \pm 0.9$	$10.0 \pm 3.8$	$118.9 \pm 13.1$	$4.3 \pm 1.8$	$10.1 \pm 3.0$	$23.9 \pm 1.6$
HD 144217	$17.3 \pm 1.6$	$6.5 \pm 1.1$	$13.5 \pm 1.5$	$25.0 \pm 2.3$	$159.3 \pm 9.1$	$5.0 \pm 2.4$	$14.0 \pm 3.6$	$50.9 \pm 1.7$
HD 145502	$33.7 \pm 1.7$	$12.2 \pm 1.2$	$14.1 \pm 2.6$	$20.5 \pm 2.5$	$199.6 \pm 8.8$	$7.8 \pm 2.0$	$30.0 \pm 2.0$	$58.8 \pm 2.5$
HD 147165	$31.3 \pm 1.6$	$16.7 \pm 1.1$	$17.5 \pm 1.1$	$26.4 \pm 2.7$	$214.2 \pm 7.7$	$10.9 \pm 2.0$	$21.1 \pm 2.0$	$61.3 \pm 2.3$
HD 147933	$57.2 \pm 5.3$	$30.6 \pm 2.6$	$17.0 \pm 2.7$	$24.9 \pm 5.0$	$173.8 \pm 16.9$	$15.5 \pm 2.8$	$28.0 \pm 3.7$	$62.5 \pm 3.6$
HD 149757	$32.6 \pm 1.6$	$14.2 \pm 1.1$	$10.3 \pm 1.2$	$16.8 \pm 2.9$	$72.0 \pm 6.9$	$10.9 \pm 2.0$	$16.7 \pm 1.9$	$46.4 \pm 2.0$
HD 164284	$13.8 \pm 1.7$	$0.4 \pm 1.3$	$6.8 \pm 1.5$	$15.7 \pm 3.0$	$111.3 \pm 9.2$	$1.8 \pm 2.0$	$11.3 \pm 2.2$	$26.9 \pm 2.7$
HD 170740	$63.3 \pm 1.8$	$24.6 \pm 1.1$	$26.3 \pm 1.2$	$52.7 \pm 2.6$	$249.6 \pm 9.9$	$20.9 \pm 1.6$	$60.7 \pm 1.7$	$122.4 \pm 2.2$
HD 198478	$75.0 \pm 2.2$	$34.6 \pm 1.6$	$33.1 \pm 1.5$	$53.3 \pm 4.2$	$379.5 \pm 11.6$	$21.2 \pm 3.5$	$46.7 \pm 4.1$	$130.6 \pm 3.4$
HD 202904	$5.7 \pm 2.3$	$1.9 \pm 1.7$	$3.6 \pm 1.8$	$15.2 \pm 3.1$	$82.2 \pm 10.6$	$3.0 \pm 2.6$	$11.7 \pm 3.5$	$18.4 \pm 2.7$
HD 207198	$132.6 \pm 1.1$	$61.1 \pm 0.7$	$32.3 \pm 1.0$	$43.2 \pm 1.7$	$227.2 \pm 9.6$	$30.0 \pm 1.8$	$71.8 \pm 2.1$	$121.8 \pm 1.9$
HD 209975	$71.5 \pm 1.4$	$26.5 \pm 1.6$	$26.9 \pm 4.5$	$43.1 \pm 3.1$	$240.2 \pm 10.0$	$25.5 \pm 2.7$	$45.5 \pm 2.6$	$114.1 \pm 3.1$
HD 214680	$20.1 \pm 0.9$	$3.9 \pm 0.9$	$5.4 \pm 1.0$	$9.2 \pm 1.6$	$68.7 \pm 7.9$	$6.4 \pm 1.5$	$4.5 \pm 1.4$	$16.1 \pm 2.0$
HD 214993	$13.6 \pm 1.3$	$0.9 \pm 0.7$	$7.6 \pm 1.4$	$10.0 \pm 2.2$	$107.1 \pm 10.0$	$4.4 \pm 1.7$	$13.9 \pm 1.7$	$18.0 \pm 2.3$
HD 218376	$38.7 \pm 1.3$	$17.2 \pm 1.0$	$14.2 \pm 1.2$	$31.6 \pm 2.3$	$175.7 \pm 10.0$	$11.2 \pm 2.0$	$37.0 \pm 2.2$	$66.0 \pm 2.2$

NOTE—All measurements in mÅ.

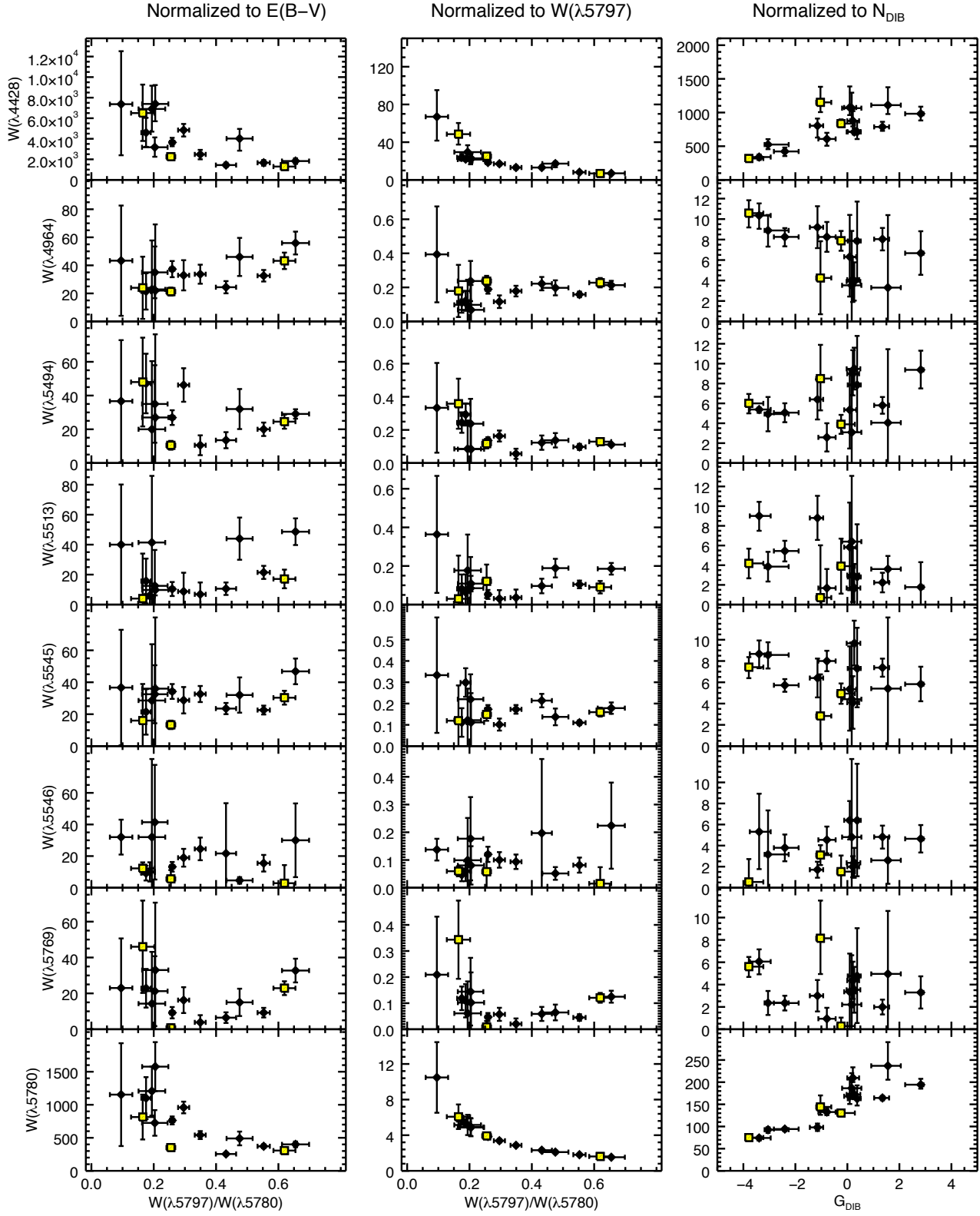
$$\begin{aligned}
PC_2 = & 0.245(z_1) - 0.137(z_2) + 0.085(z_3) - 0.185(z_4) - 0.058(z_5) - 0.203(z_6) - 0.238(z_7) \\
& + 0.306(z_8) - 0.020(z_9) - 0.085(z_{10}) + 0.168(z_{11}) + 0.264(z_{12}) + 0.361(z_{13}) + 0.077(z_{14}) \\
& - 0.008(z_{15}) + 0.145(z_{16}) - 0.055(z_{17}) + 0.109(z_{18}) - 0.226(z_{19}) - 0.015(z_{20}) - 0.334(z_{21}) \\
& - 0.287(z_{22}) - 0.403(z_{23}) \\
= & (6.98 \times 10^{-4})W(\lambda 4428) - (2.40 \times 10^{-2})W(\lambda 4964) + (2.06 \times 10^{-2})W(\lambda 5494) - (4.77 \times 10^{-2})W(\lambda 5513) \\
& - (1.35 \times 10^{-2})W(\lambda 5545) + (9.21 \times 10^{-2})W(\lambda 5546) - (8.53 \times 10^{-2})W(\lambda 5769) + (3.97 \times 10^{-3})W(\lambda 5780) \\
& - (7.27 \times 10^{-4})W(\lambda 5797) - (6.23 \times 10^{-3})W(\lambda 5850) + (2.02 \times 10^{-2})W(\lambda 6196) + (1.89 \times 10^{-2})W(\lambda 6270) \\
& + (4.29 \times 10^{-3})W(\lambda 6284) + (1.04 \times 10^{-2})W(\lambda 6376) - (4.29 \times 10^{-4})W(\lambda 6379) + (4.11 \times 10^{-3})W(\lambda 6614) \\
& - (4.13 \times 10^{-1})E(B - V) + (1.19 \times 10^{-22})N(H_1) - (8.69 \times 10^{-22})N(H_2) - (1.23 \times 10^{-23})N(H) \\
& - (1.43)f(H_2) - (1.34)F_\star - (2.57)W(\lambda 5797)/W(\lambda 5780) - 0.706
\end{aligned} \tag{A2}$$

$$\begin{aligned}
PC_3 = & 0.033(z_1) + 0.138(z_2) + 0.036(z_3) + 0.287(z_4) - 0.024(z_5) + 0.178(z_6) + 0.396(z_7) \\
& - 0.022(z_8) + 0.008(z_9) + 0.047(z_{10}) - 0.132(z_{11}) - 0.162(z_{12}) - 0.131(z_{13}) - 0.025(z_{14}) \\
& - 0.144(z_{15}) - 0.202(z_{16}) - 0.161(z_{17}) + 0.435(z_{18}) - 0.263(z_{19}) + 0.221(z_{20}) - 0.472(z_{21}) \\
& - 0.141(z_{22}) - 0.082(z_{23}) \\
= & (3.82 \times 10^{-4})W(\lambda 4428) + (6.07 \times 10^{-3})W(\lambda 4964) - (5.74 \times 10^{-2})W(\lambda 5494) + (1.08 \times 10^{-2})W(\lambda 5513) \\
& - (3.17 \times 10^{-2})W(\lambda 5545) - (1.66 \times 10^{-2})W(\lambda 5546) + (3.77 \times 10^{-4})W(\lambda 5769) - (2.84 \times 10^{-4})W(\lambda 5780) \\
& + (3.06 \times 10^{-4})W(\lambda 5797) + (3.50 \times 10^{-3})W(\lambda 5850) - (1.58 \times 10^{-2})W(\lambda 6196) - (1.16 \times 10^{-2})W(\lambda 6270) \\
& - (1.56 \times 10^{-3})W(\lambda 6284) - (3.42 \times 10^{-3})W(\lambda 6376) - (7.96 \times 10^{-3})W(\lambda 6379) - (5.73 \times 10^{-3})W(\lambda 6614) \\
& - (1.22)E(B - V) + (4.74 \times 10^{-22})N(H_1) - (1.01 \times 10^{-21})N(H_2) + (1.86 \times 10^{-22})N(H) \\
& - 2.02f(H_2) + (6.60 \times 10^{-1})F_\star + (5.22 \times 10^{-1})W(\lambda 5797)/W(\lambda 5780) + 0.99
\end{aligned} \tag{A3}$$

$$\begin{aligned}
PC_4 = & 0.134(z_1) + 0.035(z_2) - 0.237(z_3) + 0.042(z_4) - 0.137(z_5) - 0.037(z_6) + 0.001(z_7) \\
& + 0.115(z_8) - 0.230(z_9) - 0.145(z_{10}) - 0.017(z_{11}) + 0.006(z_{12}) + 0.085(z_{13}) - 0.169(z_{14}) \\
& - 0.179(z_{15}) - 0.030(z_{16}) + 0.200(z_{17}) + 0.238(z_{18}) + 0.109(z_{19}) + 0.230(z_{20}) + 0.126(z_{21}) \\
& + 0.644(z_{22}) - 0.394(z_{23}) \\
= & (3.82 \times 10^{-4})W(\lambda 4428) + (6.07 \times 10^{-3})W(\lambda 4964) - (5.75 \times 10^{-2})W(\lambda 5494) + (1.08 \times 10^{-2})W(\lambda 5513) \\
& - (3.17 \times 10^{-2})W(\lambda 5545) - (1.66 \times 10^{-2})W(\lambda 5546) + (3.77 \times 10^{-4})W(\lambda 5769) + (1.49 \times 10^{-3})W(\lambda 5780) \\
& - (8.27 \times 10^{-3})W(\lambda 5797) - (1.07 \times 10^{-2})W(\lambda 5850) - (2.06 \times 10^{-3})W(\lambda 6196) + (4.50 \times 10^{-4})W(\lambda 6270) \\
& + (1.01 \times 10^{-3})W(\lambda 6284) - (2.28 \times 10^{-2})W(\lambda 6376) - (9.92 \times 10^{-3})W(\lambda 6379) - (8.51 \times 10^{-4})W(\lambda 6614) \\
& + 1.51E(B - V) + (2.59 \times 10^{-22})N(H_1) + (4.19 \times 10^{-22})N(H_2) + (1.93 \times 10^{-22})N(H) \\
& + (5.37 \times 10^{-1})f(H_2) + 3.01F_\star - (2.51 \times 10^{-1})W(\lambda 5797)/W(\lambda 5780) - 1.84
\end{aligned} \tag{A4}$$

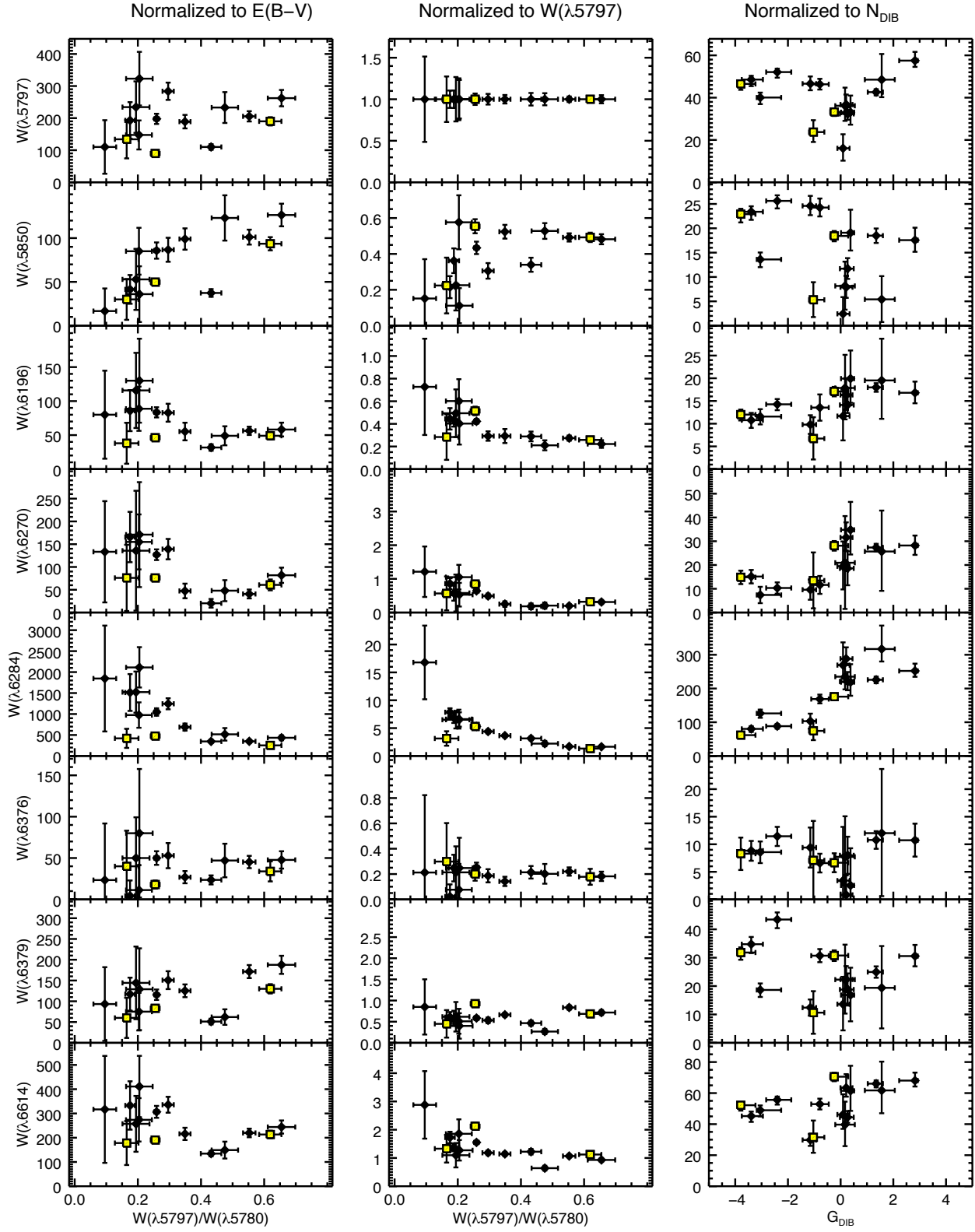
### C. ADDITIONAL FIGURES

As explained in the text and illustrated for 3 DIBs in Fig. 16, we can use different normalizations to better show the environmental responses of the DIBs. Below are similar figures, but for the remaining DIBs.



**Figure 17.** (Left:) DIBs normalized to E(B-V) as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . (Centre:) DIBs normalized to  $W(\lambda 5797)$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . (Right:) DIBs normalized to  $N_{DIB}$  as a function of  $G_{DIB}$ . Be stars are indicated by yellow squares.  $\lambda 5513$  for HD 40111, and  $\lambda 5513$  for both HD 15137 and HD 35149 have best value EW measurements equal to zero. Since fractional uncertainties are undefined, these data points have been excluded from the plots. HD 143275 has been excluded from the left plots because it has an E(B-V) value of zero.





**Figure 17.** (Continued) (*Left:*) DIBs normalized to  $E(B-V)$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . (*Centre:*) DIBs normalized to  $W(\lambda 5797)$  as a function of  $W(\lambda 5797)/W(\lambda 5780)$ . (*Right:*) DIBs normalized to  $N_{DIB}$  as a function of  $G_{DIB}$ . Be stars are indicated by yellow squares. HD 143275 has been excluded from the left plots because it has an  $E(B-V)$  value of zero.

## REFERENCES

- Abdi, H., & Williams, L. J. 2010, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433
- Adams, W. S. 1949, *ApJ*, 109, 354
- Adamson, A. J., Whittet, D. C. B., & Duley, W. W. 1991, *MNRAS*, 252, 234
- Andersson, B.-G., & Potter, S. B. 2007, *ApJ*, 665, 369
- Baron, D., Poznanski, D., Watson, D., Yao, Y., & Prochaska, J. X. 2015, *MNRAS*, 447, 545
- Bhatt, N. H., & Cami, J. 2015, *ApJS*, 216, 22
- Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, *ApJ*, 224, 132
- Bondar, A. 2012, *MNRAS*, 423, 725
- Budavári, T., Wild, V., Szalay, A. S., Dobos, L., & Yip, C.-W. 2009, *MNRAS*, 394, 1496
- Cami, J., Sonnentrucker, P., Ehrenfreund, P., & Foing, B. H. 1997, *A&A*, 326, 822
- Campbell, E. K., Holz, M., Gerlich, D., & Maier, J. P. 2015, *Nature*, 523, 322
- Carrasco, L., Strom, S. E., & Strom, K. M. 1973, *ApJ*, 182, 95
- Conselice, C. J. 2006, *MNRAS*, 373, 1389
- Cox, N. L. J., Cordiner, M. A., Cami, J., et al. 2006, *A&A*, 447, 991
- Dahlstrom, J., York, D. G., Welty, D. E., et al. 2013, *ApJ*, 773, 41
- Dekker, H., D’Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Iye & A. F. Moorwood, 534–545
- Donati, J.-F. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 307, Solar Polarization, ed. J. Trujillo-Bueno & J. Sanchez Almeida, 41
- Federman, S. R., Vanden Bout, P. A., & Kumar, C. K. 1984, *ApJ*, 282, 485
- Fitzpatrick, E. L., & Massa, D. 2007, *ApJ*, 663, 320
- Friedman, S. D., York, D. G., McCall, B. J., et al. 2011, *ApJ*, 727, 33
- Galazutdinov, G. A., Krelowski, J., Moutou, C., & Musaev, F. A. 1998, *MNRAS*, 295, 437
- Gnaniński, P., Sikorski, J. K., & Galazutdinov, G. A. 2007, *A&A*, 469, 201
- Gould, R. J., & Salpeter, E. E. 1963, *ApJ*, 138, 393
- Habing, H. J. 1968, *BAN*, 19, 421
- Heger, M. L. 1922, *Lick Observatory Bulletin*, 10, 146
- Herbig, G. H. 1968, *ZA*, 68, 243
- . 1975, *ApJ*, 196, 129
- . 1993, *ApJ*, 407, 142
- . 1995, *ARA&A*, 33, 19
- Hobbs, L. M., York, D. G., Snow, T. P., et al. 2008, *ApJ*, 680, 1256
- Hobbs, L. M., York, D. G., Thorburn, J. A., et al. 2009, *ApJ*, 705, 32
- Hotelling, H. 1933, *Journal of educational psychology*, 24, 417
- Jenkins, E. B. 2009, *ApJ*, 700, 1299
- Jenniskens, P., Ehrenfreund, P., & Foing, B. 1994, *A&A*, 281, 517
- Jolliffe, I. 2014, *Principal Component Analysis* (John Wiley & Sons, Ltd)
- Jura, M. 1980, *ApJ*, 235, 63
- Kausch, W., Noll, S., Smette, A., et al. 2015, *A&A*, 576, A78
- Kos, J., & Zwitter, T. 2013, *ApJ*, 774, 72
- Krelowski, J., Ehrenfreund, P., Foing, B. H., et al. 1999, *A&A*, 347, 235
- Krelowski, J., Schmidt, M., & Snow, T. P. 1997, *PASP*, 109, 1135
- Krelowski, J., & Sneden, C. 1995, in *Astrophysics and Space Science Library*, Vol. 202, The Diffuse Interstellar Bands, ed. A. G. G. M. Tielens & T. P. Snow, 13
- Krelowski, J., & Walker, G. A. H. 1987, *ApJ*, 312, 860
- Kumar, C. K. 1986, *ApJ*, 306, 38
- Kumar, C. K., Federman, S. R., & Vanden Bout, P. A. 1982, *ApJL*, 261, L51
- Lan, T.-W., Ménard, B., & Zhu, G. 2015, *MNRAS*, 452, 3629
- Le Page, V., Snow, T. P., & Bierbaum, V. M. 2001, *ApJS*, 132, 233
- . 2003, *ApJ*, 584, 316
- Markwardt, C. B. 2009, in *Astronomical Society of the Pacific Conference Series*, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251
- McCall, B. J., Drosback, M. M., Thorburn, J. A., et al. 2010, *ApJ*, 708, 1628
- Moultaka, J., Ilovaisky, S. A., Prugniel, P., & Soubiran, C. 2004, *PASP*, 116, 693
- Oka, T., Welty, D. E., Johnson, S., et al. 2013, *ApJ*, 773, 42
- Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, *A&A*, 530, A50
- Pearson. 1901, *Philosophical Magazine Series 6*, 2, 559
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2001, *ApJ*, 555, 839
- Sarre, P. J. 2006, *Journal of Molecular Spectroscopy*, 238, 1
- Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, *ApJ*, 216, 291
- Schild, R. E. 1978, *ApJS*, 37, 77

- Seab, C. G., & Snow, Jr., T. P. 1984, *ApJ*, 277, 200
- Sigut, T. A. A., & Patel, P. 2013, *ApJ*, 765, 41
- Smette, A., Sana, H., Noll, S., et al. 2015, *A&A*, 576, A77
- Snow, T. P. 2014, in *IAU Symposium*, Vol. 297, IAU Symposium, ed. J. Cami & N. L. J. Cox, 3–12
- Snow, T. P., & McCall, B. J. 2006, *ARA&A*, 44, 367
- Snow, T. P., Zukowski, D., & Massey, P. 2002, *ApJ*, 578, 877
- Snow, Jr., T. P. 1973, *PASP*, 85, 590
- Snow, Jr., T. P., & Wallerstein, G. 1972, *PASP*, 84, 492
- Sonnentrucker, P., Cami, J., Ehrenfreund, P., & Foing, B. H. 1997, *A&A*, 327, 1215
- Sonnentrucker, P., Welty, D. E., Thorburn, J. A., & York, D. G. 2007, *ApJS*, 168, 58
- Suzuki, N. 2006, *ApJS*, 163, 110
- Thorburn, J. A., Hobbs, L. M., McCall, B. J., et al. 2003, *ApJ*, 584, 339
- Tielens, A. G. G. M., & Hollenbach, D. 1985, *ApJ*, 291, 722
- van Dishoeck, E. F., & Black, J. H. 1982, *ApJ*, 258, 533
- Vos, D. A. I., Cox, N. L. J., Kaper, L., Spaans, M., & Ehrenfreund, P. 2011, *A&A*, 533, A129
- Vuong, M. H., & Foing, B. H. 2000, *A&A*, 363, L5
- Walker, G. A. H., Bohlender, D. A., Maier, J. P., & Campbell, E. K. 2015, *ApJL*, 812, L8
- Wegner, W. 2003, *Astronomische Nachrichten*, 324, 219
- Welty, D. E., Federman, S. R., Gredel, R., Thorburn, J. A., & Lambert, D. L. 2006, *ApJS*, 165, 138
- Welty, D. E., & Hobbs, L. M. 2001, *ApJS*, 133, 345
- Welty, D. E., Hobbs, L. M., & Kulkarni, V. P. 1994, *ApJ*, 436, 152
- Williams, M. J., Bureau, M., & Cappellari, M. 2010, *MNRAS*, 409, 1330
- Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004, *AJ*, 128, 2603